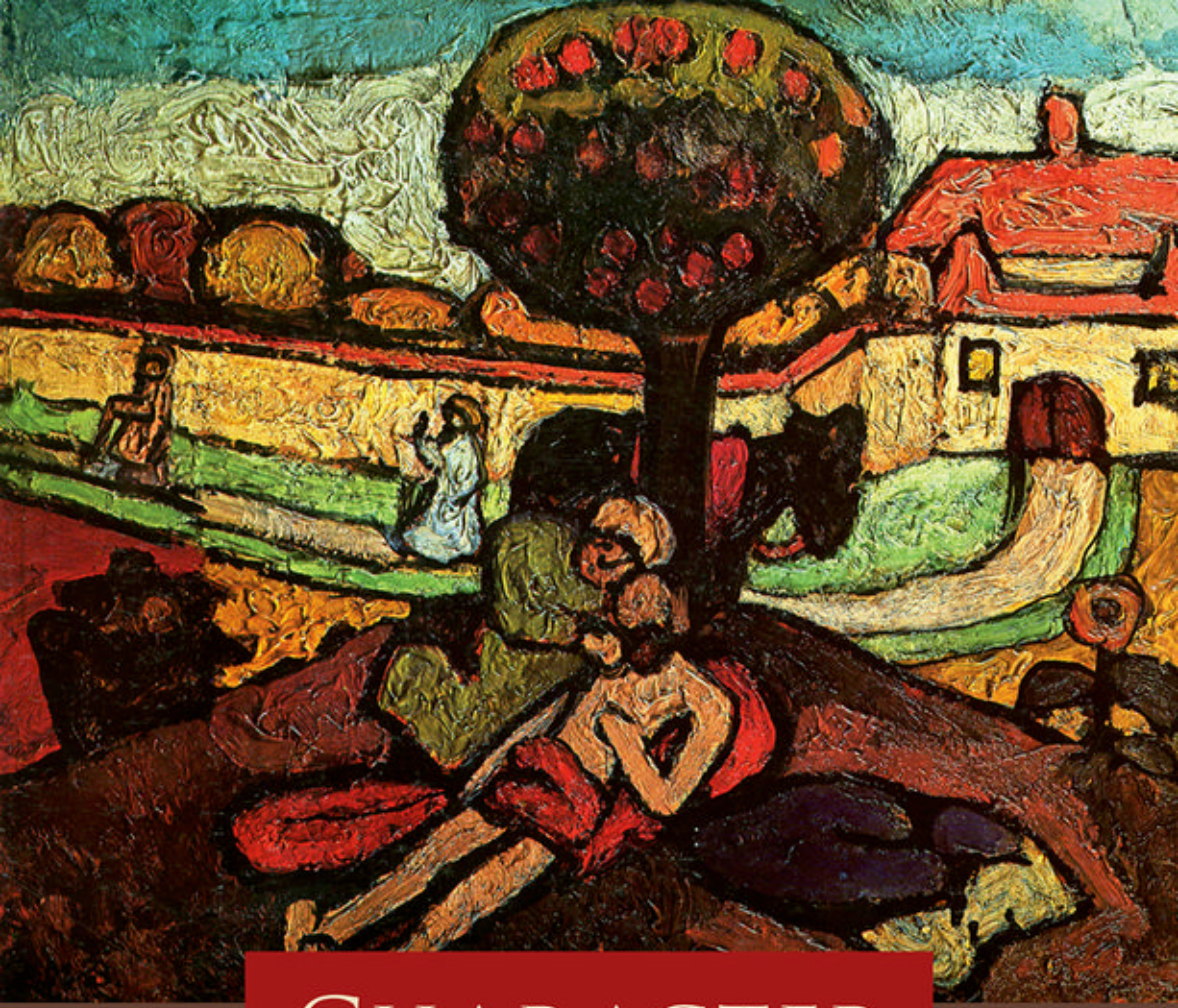

Character

New Directions from Philosophy, Psychology, and Theology

Edited by

CHRISTIAN B. MILLER

R. MICHAEL FURR

ANGELA KNOBEL

WILLIAM FLEESON

Character

Character

NEW DIRECTIONS FROM PHILOSOPHY, PSYCHOLOGY, AND THEOLOGY

Edited by Christian B. Miller, R. Michael Furr, Angela Knobel, William Fleeson

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of
Oxford. It furthers the University's objective of excellence in research,
scholarship, and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

© Oxford University Press 2015

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the prior
permission in writing of Oxford University Press, or as expressly permitted by law,
by license, or under terms agreed with the appropriate reproduction rights organization.
Inquiries concerning reproduction outside the scope of the above should be sent to the
Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
Character : new directions from philosophy, psychology, and theology / edited by
Christian B. Miller, R. Michael Furr, Angela Knobel, William Fleeson.
pages cm
Includes index.
ISBN 978–0–19–020460–0 (cloth : alk. paper) 1. Character. I. Miller, Christian B.
II. Furr, R. Michael. III. Knobel, Angela. IV. Fleeson, William.
BJ1518.C43 2015
170—dc23
2014046054

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

TABLE OF CONTENTS

Contributors ix

Introduction 1
CHRISTIAN B. MILLER, R. MICHAEL FURR, ANGELA KNOBEL, WILLIAM FLEESON

PART I } **Overview of the Study of Character in Philosophy and Psychology**

1. **Some Foundational Questions in Philosophy about Character** 19
CHRISTIAN B. MILLER AND ANGELA KNOBEL

2. **Personality Science and the Foundations of Character** 41
WILLIAM FLEESON, R. MICHAEL FURR, ERANDA JAYAWICKREME, ERIK G. HELZER, ANSELMA G. HARTLEY, AND PETER MEINDL

PART II } **Beliefs about Character**

3. **Lay Beliefs in True Altruism versus Universal Egoism** 75
JOCHEN E. GEBAUER, CONSTANTINE SEDIKIDES, MARK R. LEARY, AND JENS B. ASENDORPF

4. **Understanding the Importance and Perceived Structure of Moral Character** 100
GEOFFREY P. GOODWIN, JARED PIAZZA, AND PAUL ROZIN

PART III } **The Existence and Nature of Character**

5. **Moving Character beyond the Person-Situation Debate: The Stable and Dynamic Nature of Virtues in Everyday Life** 129
WIEBKE BLEIDORN

6. **Character Traits in the Workplace: A Three-Month Diary Study of Moral and Immoral Organizational Behaviors** 150
TAYA R. COHEN AND A. T. PANTER

7. The Mixed Trait Model of Character Traits and the Moral Domains of Resource Distribution and Theft 164
CHRISTIAN B. MILLER

8. Emotion and Character 192
CHARLES STARKEY

PART IV } Character and Ethical Theory

9. Taking Moral Risks and Becoming Virtuous 215
REBECCA STANGL

10. Dispositions, Character, and the Value of Acts 233
BRADFORD COKELET

11. Exemplarism and Admiration 251
LINDA ZAGZEBSKI

PART V } Virtue Epistemology

12. People Listen to People Who Listen: Instilling Virtues of Deference 271
KRISTOFFER AHLSTROM-VIJ

13. "Why Can't We Be Friends?": Reflections on Empirical Psychology and Virtue Epistemology 288
NATHAN L. KING

14. From Virtue Epistemology to Abilism: Theoretical and Empirical Developments 315
JOHN TURRI

PART VI } Particular Virtues

15. Christian Humility as a Social Virtue 333
MICHAEL W. AUSTIN

16. A Different Kind of Wisdom 351
ANGELA KNOBEL

17. Bearing Burdens and the Character of God in the Hebrew Bible 368
CRISTIAN MIHUT

18. Domain Specificity in Self-Control 393
ANGELA LEE DUCKWORTH AND ELI TSUKAYAMA

19. **Can Text Messages Make People Kinder?** 412
SARA KONRATH

PART VII } **Character Development**

20. **The Emergence of Moral Character in Infancy: Developmental Changes and Individual Differences in Fairness Concerns and Prosocial Behavior during the First Two Years of Life** 445
JESSICA A. SOMMERVILLE

21. **Character Development in the School Years: Relations among Theory of Mind, Moral Identity, and Positive and Negative Behavior toward Peers** 467
ELIZABETH A. BOERGER AND ANTHONY J. HOFFMAN

22. **Character across Early Emerging Adulthood: Character Traits, Character Strivings, and Moral Self-Attributes** 490
ERIK E. NOFTLE

23. **Etiquette and Exemplarity in Judaism** 522
TZVI NOVICK

24. **Christian Character Formation and the Infusion of Grace** 538
RAY S. YEO

25. **Necessity and Human Agency: Cultivating Character in the Reformed Christian Tradition** 555
ELIZABETH AGNEW COCHRAN

26. **Liturgy and the Moral Life** 572
TERENCE D. CUNEO

27. **Cultivating Virtues through Sartorial Practices: The Case of the Islamic Veil in Indonesia** 590
ELIZABETH M. BUCAR

PART VIII } **Challenges to Character and Virtue from Neuroscience and Situationism**

28. **Character Traits and the Neuroscience of Social Behavior** 605
DANIEL J. McKAUGHAN

29. **Character and Coherence: Testing the Stability of Naturalistically Observed Daily Moral Behavior** 630
MATTHIAS R. MEHL, KATHRYN L. BOLLICH, JOHN M. DORIS, AND SIMINE VAZIRE

30. **Taking Evil into the Lab: Exploring the Frontiers of Morality and Individual Differences** 652
DAVID GALLARDO-PUJOL, ELIZABET OREKHOVA, VERÓNICA BENET-MARTÍNEZ, AND MEL SLATER

31. **War Crimes: Causes, Excuses, and Blame** 671
MATTHEW TALBERT AND JESSICA WOLFENDALE

Index 693

CONTRIBUTORS

Kristoffer Ahlstrom-Vij is Senior Lecturer in Philosophy at the University of Kent.

Jens B. Asendorpf is Professor Emeritus of Personality Psychology at the Humboldt-Universität zu Berlin.

Mike Austin is Professor of Philosophy at Eastern Kentucky University.

Verónica Benet-Martínez is ICREA Research Professor in Social and Behavioral Sciences at the Pompeu Fabra University

Wiebke Bleidorn is Assistant Professor of Psychology at University of California, Davis.

Elizabeth A. Boerger is Assistant Professor of Psychology at Slippery Rock University.

Kathryn L. Bollich is a Graduate Student in the Psychology Department at Washington University in St. Louis.

Elizabeth Bucar is Associate Professor of Religious Studies at Northeastern University.

Elizabeth Agnew Cochran is Associate Professor of Theology at Duquesne University.

Taya R. Cohen is Assistant Professor of Organizational Behavior and Theory at the Tepper School of Business, Carnegie Mellon University.

Bradford Cokelet is Assistant Professor of Philosophy at the University of Miami.

Terence Cuneo is Marsh Professor of Intellectual and Moral Philosophy at the University of Vermont.

John M. Doris is Professor of Philosophy and Philosophy-Neuroscience-Psychology at the Washington University in St. Louis.

Angela Lee Duckworth is Associate Professor of Psychology at the University of Pennsylvania.

William Fleeson is Professor of Psychology at Wake Forest University.

R. Michael Furr is Professor of Psychology at Wake Forest University.

David Gallardo-Pujol is Senior Lecturer in Individual Differences at the University of Barcelona.

Jochen Gebauer is an Emmy Noether Junior Research Group leader at the University of Mannheim.

Geoffrey P. Goodwin is Associate Professor of Psychology at the University of Pennsylvania.

Anselma G. Hartley is a Postdoctoral Research Associate in Psychology at Wake Forest University.

Erik G. Helzer is Assistant Professor at the Johns Hopkins Carey Business School.

Anthony J Hoffman is an Undergraduate Student at Slippery Rock University.

Eranda Jayawickreme is Assistant Professor of Psychology at Wake Forest University.

Nathan King is Assistant Professor of Philosophy at Whitworth University.

Angela Knobel is Associate Professor of Philosophy at The Catholic University of America.

Sara Konrath is Assistant Professor of Philanthropic Studies at the Lilly Family School of Philanthropy at Indiana University.

Mark R. Leary is Professor of Psychology and Neuroscience at Duke University.

Daniel J. McKaughan is Associate Professor of Philosophy at Boston College.

Matthias R. Mehl is Associate Professor of Psychology at the University of Arizona.

Peter Meindl is a Graduate Student in the Psychology Department at the University of Southern California.

Cristian Mihut is Associate Professor of Philosophy at Bethel College.

Christian B. Miller is Professor of Philosophy at Wake Forest University.

Erik E. Noftle is Assistant Professor of Psychology at Willamette University.

Tzvi Novick is Associate Professor of Theology at University of Notre Dame.

Elizabet Orekhova is a PhD student in the Department of Personality, Assessment, and Psychological Treatment at the University of Barcelona.

A. T. Panter is Senior Associate Dean for Undergraduate Education and Professor of Psychology in the L. L. Thurstone Psychometric Laboratory at University of North Carolina at Chapel Hill.

Jared Piazza is a Lecturer in Psychology at Lancaster University.

Paul Rozin is Professor of Psychology at the University of Pennsylvania.

Constantine Sedikides is Professor of Social and Personality Psychology at the University of Southampton.

Clemens Sedmak is FD Maurice Professor of Moral and Social Theology at King's College, London.

Mel Slater is ICREA Research Professor at the Department of Personality, Evaluation, and Psychological Treatment at the University of Barcelona.

Jessica A. Sommerville is Associate Professor of Psychology at University of Washington.

Rebecca Stangl is Associate Professor of Philosophy at the University of Virginia.

Charles Starkey is Associate Professor of Philosophy and Fellow of the Rutland Institute for Ethics at Clemson University.

Matthew Talbert is Associate Professor of Philosophy at West Virginia University.

Eli Tsukayama is a Post-Doctoral Fellow in the Department of Psychology at University of Pennsylvania.

John Turri is Associate Professor of Philosophy and a member of the Cognitive Science Program at University of Waterloo.

Simine Vazire is Associate Professor of Psychology at University of California, Davis.

Jessica Wolfendale is Associate Professor of Philosophy at West Virginia University.

Ray Yeo is an Independent Scholar

Linda Zagzebski is the George Lynn Cross Research Professor and Kingfisher College Chair of the Philosophy of Religion and Ethics at the University of Oklahoma.

Character

Introduction

Christian B. Miller, R. Michael Furr, Angela Knobel, and William Fleeson

We are very proud to present the diverse, creative, and important work reflected in the 31 chapters of this volume. These papers stem from the work of the Character Project (www.thecharacterproject.com) at Wake Forest University. The Character Project has been dedicated to addressing fundamental questions regarding, among other things, the existence and nature of character traits such as honesty and compassion, the role of such traits in ethical theory, the potential for improving our character and overcoming our character flaws, and the possibility that thinking about human and divine character might be central to theological ethics. Generously supported by the John Templeton Foundation, the Character Project completed three years of activities ranging from multiple conferences to the creation of new websites to psychology studies to a prize competition.

Central to this work were three funding competitions on the psychology, philosophy, and theology of character. These competitions were especially aimed at identifying young scholars ten years or less out of graduate school who were doing innovative and exciting new work on character. The response we received to these competitions was overwhelming. There were 128 proposals submitted in psychology, 110 in philosophy, and 60 in theology. After a lengthy process involving numerous internal and external reviewers, grants were awarded to 28 scholars from the United States and Europe totaling $3 million. These scholars presented their central conclusions at two conferences held at Wake Forest University. These presentations served as the basis for the chapters in this collection.

Hence this volume showcases cutting edge research by some of the best scholars in these three disciplines, whose work has not only passed through a rigorous initial funding competition but has also been strengthened by the feedback and criticisms of leading interdisciplinary scholars working on

character today. The volume also contains four chapters by the editors themselves (chapters 1, 2, 7, and 16).

Unlike most edited collections where the editors invite already established figures on the relevant topic, this collection is truly unique in featuring the work of scholars who emerged out of 298 applicants. In addition, most of these scholars are young, up-and-coming voices in their fields with new perspectives to offer. We hope that this volume will constitute some of the most novel and groundbreaking work done on character in many years.

In this brief introduction, we set the stage for the chapters to come by outlining some of the questions which are front and center in contemporary thinking about character in the fields of psychology, philosophy, and theology. We then offer detailed summaries of each of the chapters as well as acknowledge the many people who have helped us over the years.

Psychology and Character

Character was an important focus of study in the early days of empirical psychology. Empirical attention to character faded quickly, though, following findings that were widely interpreted as evidence that broad character traits, such as compassion and honesty, did not exist. By the late 1920s, the study of "character" in psychology shifted to the study of "personality," and personality psychology has since remained one of the core subdisciplines within psychology. With a focus on understanding behavioral differences among people, personality psychology shares much in common with the early study of character. However, the shift from character to personality was more than a rebranding. Perhaps partly to avoid any reminder of "character" in their studies, personality psychologists since the 1920s had generally abandoned the study of ethics and morality.

In the 1960s, questions about the existence of psychological traits were raised again, this time questioning the existence of not only character but also of personality more generally. Resonating with the earlier findings that had led psychological science away from character, this later research seemingly indicated that behaviors on single occasions were surprisingly inconsistent—people who behaved in a particular way in one situation were not likely to behave that way in another situation. In addition, experimental studies of the effects of situations produced surprising and provocative results, suggesting that behavior is not primarily the product of broad character traits, such as honesty or compassion, but rather of the situations in which people find themselves. In other words, evidence suggested that researchers should look, not to internal features of individuals, but, rather, to external facts about situations in order to explain why people behave the way they do.

This "situationist" view of behavior dominated for two or three decades, and it led many psychologists to conclude that one's personality characteristics (e.g., traits, dispositions, values, etc.) were, at best, only weakly connected to one's behavior. In fact, it lead more than a few to conclude that personality itself was little more than a fiction that people create in their struggle to makes sense of themselves and other people.

More recent research reveals that such extreme situationist positions are no longer tenable. Indeed, the past 20 to 30 years have seen significant demonstrations of the meaningfulness and empirical viability of stable personality dispositions. Researchers have found considerable empirical support for the stability of traits over time, as well as for robust predictions between traits and important outcomes such as marital quality, mental health, and longevity. Researchers have also achieved new insights into the nature and magnitude of consistency in behavior, and they have made significant progress in resolving the paradox of both consistency and inconsistency in behavior existing simultaneously. Most psychologists no longer doubt the existence of personality characteristics such as extraversion or emotional stability.

Despite this progress in understanding personality in general, there has been little such progress in understanding the personological basis of moral character. Thus, there are many fundamental unanswered questions, including the following:

1. Are there reliable and meaningful individual differences in dispositions to think, feel, and act in morally relevant ways?
2. Is there evidence, including consistency evidence, for the existence of character traits?
3. If character traits do stand up to empirical scrutiny, are they best conceptualized as general traits, such as those in the Big Five approach (openness, conscientiousness, extraversion, agreeableness, and neuroticism), or are they better understood as hundreds of highly localized traits? If they are general traits, does the Big Five approach capture all of them?
4. What are the empirically most promising psychological models of character traits, including integrating both character traits and situations in the explanation and prediction of behavior?
5. What roles do beliefs, desires, and emotions play in character traits?
6. Do goals and identities have important relationships to character?
7. How much self-regulation do people have over their character traits at a given moment and over a given period of time?
8. What is the best way to define situations in the first place in order to empirically assess the role of situations in influencing character-relevant behavior?
9. How does character develop and how it is perceived?

10. How accurate is the understanding people typically have of their own character traits? And how accurate is the understanding people have of other people's character traits?
11. How do people acquire character traits over time? Are certain strategies more effective in cultivating a given character trait than others?
12. What new (or old) methods and analytic techniques can best be utilized to reach breakthroughs in the study of character?

The chapters in this volume written by psychologists—specifically chapters 2 (Fleeson and Furr et al.), 3 (Gebauer et al.), 4 (Goodwin et al.), 5 (Bleidorn), 6 (Cohen and Panter), 18 (Duckworth and Tsukayama), 19 (Konrath), 20 (Sommerville), 21 (Boerger and Hoffman), 22 (Noftle), 29 (Mehl et al.), and 30 (Gallardo-Pujol et al.)—cumulatively address all these questions and make important progress in trying to answer them.

Philosophy and Character

Turning to work on character in philosophy: it is widely held that twentieth-century thinking about ethics largely ignored notions of character and virtue, focusing instead on rules and on the consequences of actions. This notable omission changed in the 1980s, however, with the rebirth of virtue ethics. A central claim made by many virtue ethicists is that ethical theorizing should begin by thinking about character, and, perhaps more specifically, about what a person with the virtues is like or what he or she would do in a given set of circumstances. Today, virtue ethics has established itself as a leading contender in contemporary ethical thought, and alternative ethical theories typically seek ways to incorporate character traits into their approaches.

At the same time, in large part due to the work of Gilbert Harman and John Doris, many philosophers have recently begun to appreciate the implications of experimental results in psychology for their thinking about character. In particular, Harman and Doris appealed to the situationist tradition in psychology in order to argue that there are no traditional character traits, such as honesty, compassion, and courage, or at least that their possession is rare. This claim has been met with a number of responses, and examining the philosophical implications of empirical work on character in general and virtue in particular has become one of the most exciting areas in contemporary philosophy.

Today there are also a number of central questions about the philosophy of character, including the following:

1. What is the best conceptual understanding of the virtues and vices, and what role do beliefs, desires, and emotions have to play in a given virtue or vice?

2. Is an ethical theory based on the virtues overly self-centered, as is sometimes alleged about Aristotelian virtue ethics?
3. Can the most plausible versions of virtue ethics be given an objective foundation, or do they fall prey to moral relativism?
4. Is virtue ethics really a legitimate alternative ethical theory to, for instance, the leading forms of consequentialism, which typically base the moral status of an action on the consequences produced by the action, or the leading forms of Kantian ethics, which typically base the moral status of an action on whether it is universalizable or treats others as ends?
5. Does the topic of character have important implications in contemporary meta-ethics, perhaps in the form of advancing the debate between moral realists and anti-realists?
6. Does the topic of character have important implications in contemporary epistemology, say, in the form of developing a virtue epistemology?
7. Does the latest work in psychology that bears on the existence and nature of character traits ultimately threaten virtue ethics or, more generally, most traditional philosophical thinking about character?
8. If the empirical results suggest that most people are neither virtuous nor vicious, then how should advocates of virtue ethics think about the characters of the vast majority of people?

The chapters written primarily from a philosophical perspective—chapters 1 (Miller and Knobel), 7 (Miller), 8 (Starkey), 9 (Stangl), 10 (Cokelet), 11 (Zagzebski), 12 (Ahlstrom-Vij), 13 (King), 14 (Turri), 28 (McKaughan), and 31 (Talbert and Wolfendale)—address many of these questions and also make significant progress in trying to answer them.

Theology and Character

Theology has also witnessed a rebirth of interest in character. In the area of religious ethics, scholars have considered whether to base ethical theories on the virtues and vices rather than on divine commands or systems of rules. Stanley Hauerwas, for instance, has long argued that Christian ethics would benefit by incorporating insights from the Aristotelian virtue ethical tradition.

If the virtues should have an important role to play in religious ethics, it is not clear how to understand the contribution that both human beings and God are supposed to make in fostering them. Some Christian theologians have maintained that not only faith, hope, and love but also the moral virtues cannot be acquired by human effort but must be infused by God. On the other hand, theologians also acknowledge many cases in which non-believers seem to exhibit virtuous characters which they have acquired on their own over time, and most acknowledge the value of writings on the virtues by non-theistic

philosophers, such as Aristotle. In addition, there has been renewed interest in better understanding specific virtues from a religious perspective. For instance, much recent work has been done on the virtues of faith, hope, and charity.

Theologians and religion scholars have also been actively exploring the role of character and virtue in the practical life of the religious believer, such as in the context of worship and liturgical practice. For instance, in the Eucharist, Christians have the opportunity to develop the virtue of gratitude for the mercy of God toward them. Through fasting and other ascetical practices, they can foster the virtues of discipline and fortitude. Confession can combat pride, while worship can help instill faith and hope, and regular offerings of time and money can cultivate generosity.

Hence, there are a number of central questions about the theology of character, including the following:

1. In what ways can religious ethics benefit from philosophical thinking about character and virtue?
2. In what ways can religious ethics benefit from experimental studies of character and virtue?
3. Should theologians and religious ethicists look to secular virtue ethical approaches to inform their ethical theorizing, and if so, what benefits can they derive from such views?
4. What is the best way to understand the virtues of faith, hope, or charity or other traits of character that have been important in religious writings?
5. What is the best model of the role of God or the Holy Spirit or some other divinity in fostering character development?
6. Should Jesus or other central figures of various religious traditions be understood as fully virtuous individuals whose actions define what a morally right action is and whose lives should be objects of emulation and play a central role in religious ethics?
7. In what ways do character and virtue play a role in people's actual religious lives?

The chapters written from an explicitly religious perspective—chapters 15 (Austin), 16 (Knobel), 17 (Mihut), 23 (Novick), 24 (Yeo), 25 (Cochran), 26 (Cuneo), and 27 (Bucar)—make important contributions in addressing many of these questions.

Overview of the Chapters

In the first chapter in Part I, two of us (Miller and Knobel) provide an overview of what we consider to be some of the foundational questions in philosophy about character. We do not try to settle any of them, but rather aim to provide

some broad conceptual background on the topic of character that hopefully will be useful in approaching the chapters that follow.

The other two editors, William Fleeson and R. Michael Furr, together with additional members of the Character Project, then offer an overview of the psychology of character in their chapter. As they note, with the recent renaissance in personality research, there has been renewed interest in the study of character and its relation to personality. Using a trait-centered approach, they provide an empirical and theoretical foundation for the recent revival of character as an object of scientific psychological study. In particular, they outline five key questions for the study of moral character, review extant research on character, and discuss areas for future research.

Part II of the collection concerns, not character traits directly, but rather our beliefs about them. Jochen Gebauer and his colleagues, for instance, consider, not the question of whether altruism exists in human beings, but rather the distinct question of to what degree lay people believe in the existence of true altruism. They note that such a lay belief can be important in its own right. More precisely, their research concerned (i) the prevalence of the belief in the existence of true altruism among lay people, (ii) the relation between this belief and other relevant lay beliefs, and (iii) the relation between the altruism belief and moral character.

Geoffrey Goodwin and his colleagues investigate how important moral character is in the impressions and evaluations we form of other people in our daily lives. They address this question by reporting on recent research showing that moral character is extremely important in person perception and evaluation, and much more so than the related notion of social warmth. Moral character is also seen as particularly fundamental to identity. In their view, there are three central questions about the perception of moral character that remain to be answered: (i) is morality properly conceived of as a third dimension of person and group perception, separate from both competence and warmth? (ii) how is morality and warmth information actually obtained in ordinary social interactions? and (iii) what is the perceived structure of moral character? They take some preliminary steps in these directions.

Part III moves from beliefs to the actual existence and nature of character traits. To start, Wiebke Bleidorn engages with the situationist debate in social psychology and takes a fresh look at the degree and meaning of consistency of character traits. Specifically, she proposes (i) that there is more than one type of behavioral consistency relevant to character traits, (ii) that character traits can be described as density distributions of virtue states, and (iii) that people flexibly adapt their virtue states to their current role context. The theoretical and empirical background for these claims is discussed in view of the recent findings of her project Loving Mum and Tough Career Girl, which examined virtuous behavior in working parents' daily lives.

The Work Experiences and Character Traits project, led by Taya Cohen and A. T. Panter, investigated how moral character, personality, emotions, and treatment by managers and coworkers affected how frequently workers engaged in ethical and unethical behavior at their jobs. Over three months, they administered 14 surveys to more than 1,500 adults living in the United States, who worked in a diverse array of occupations in the private and public sectors. Their research showed that individual differences in moral character have consistent, meaningful effects on employees' work behaviors, and their findings contest situationist perspectives that de-emphasize the importance of individual differences in predicting behavior.

The Director of the Character Project, Christian Miller, expands on the Mixed Trait model he has developed for understanding the moral character traits of most people. On this model, most people have neither any moral virtues nor any moral vices but rather traits of character which are morally positive in some respects but negative in others. In developing this approach in previous work, he relied primarily on empirical work pertaining to the moral domains of helping, physical aggression, lying, and cheating. The goal of this chapter is to extend his earlier discussion to two additional areas—resource distribution and theft. In doing so, he considers whether the existing research is compatible with his Mixed Trait model, or whether instead it gives him reason to be concerned with how broadly applicable the model really is. His conclusion is that the results are, so to speak, a mixed bag.

Finally, Charles Starkey notes that despite the tremendous growth of interest in both emotion and character in recent years, little has been said about the relation between the two. He argues that emotions have a proximal and fundamental role in determining character. The proximal role consists in the effects of emotion on the way a person perceives and subsequently cognizes the object of emotion. This plays a significant part in determining character-relevant actions on his view. The fundamental role consists in the function that emotions have in sustaining values on which character traits are based. Emotions, Starkey claims, enable character traits to persist by maintaining the importance of these values and thus countering what he calls "axiological entropy," which is the diminution over time of the tacit sense of importance of such values. Emotions are thus essential to the possession of character traits, and character development involves developing an appropriate emotional repertoire.

The three chapters in Part IV focus specifically on questions about character and normative theory. The first, by Rebecca Stangl, begins by noting that many contemporary virtue theorists hold that an active life of virtue is the greatest of goods. She then claims that moral agents sometimes face what she calls a situation of "moral risk," where an agent is capable of achieving some important non-moral good for others, but only by embarking on a course of action that poses a genuine risk to her character, and so poses a genuine risk to her leading an active life of virtue. Stangl observes that if an active life of virtue

is the greatest of goods, then it might seem as if taking such risks is never permissible. But she argues that taking moral risks is not only compatible with the claim that an active life of virtue is the greatest of goods, but also that, even given that claim, taking such risks is sometimes required.

Bradford Cokelet also focuses on virtue ethics, and specifically the central virtue-ethical thesis that the ethical quality of an agent's actions is a function of her dispositional character. According to Cokelet, skeptics have distinguished (rightly, in his view) between an agent's particular intentions or occurrent motives and dispositional facts about her character. But they then contend (mistakenly, he thinks) that if we are attentive to this distinction, we will see that the virtue ethical thesis is false. In his chapter Cokelet presents a new interpretation and defense of the virtue ethical thesis and attempts to rebuff the skeptical attacks advanced by Thomas Hurka, Julia Markovits, and Roger Crisp. The key, he argues, is for virtue ethicists to adopt a personal values conception of character instead of a reliable trait conception.

The third chapter in this part on ethical theory is by Linda Zagzebski, and begins by outlining the nature of admiration, the kinds of admiration, and some psychological studies of admiration. Zagzebski then describes a way that admiration can be made foundational in a moral theory, which she calls "exemplarist virtue theory." In this theory basic moral concepts are defined via direct reference to exemplars of moral goodness, picked out through the reflective use of the emotion of admiration.

Moral traits are not the only kind of character traits, and the three chapters in Part V primarily explore epistemic traits instead. Kristoffer Ahlstrom-Vij begins by noting that we often fail to defer to sources who know what they are talking about, and when doing so consistently, we fail to manifest a virtue of deference. This is because on his view epistemic virtues are dispositions that promote epistemic goals, and knowledge is an epistemic goal. His chapter makes two points about how to instill this virtue. First, virtues of deference can be instilled by promoting compliance with requests on the part of good sources to be listened to, since listening is conducive to believing. Second, recent research in social psychology on the relationship between justice and compliance suggests that one way to promote compliance with such requests is by having the relevant sources manifest a virtue of their own—namely, that of procedural justice, which involves as a crucial component being prepared to listen.

Nathan King's focus is on virtue epistemology, an approach to epistemology that gives the notion of an intellectual virtue a fundamental role. Recently virtue epistemology has been challenged by advocates of epistemic situationism, the view (as he understands it) that human cognition typically results not from intellectual virtue, but rather is highly susceptible to epistemically irrelevant factors in a subject's environment. Crucially, say situationist critics, empirical psychology reveals that the intellectual character traits and faculties

most human thinkers have are not the virtues that figure prominently in virtue epistemologies. King argues that once the varieties of epistemic situationism and virtue epistemology are properly understood, the appearance of conflict largely dissipates.

Finally, John Turri reviews several theoretical and empirical developments relevant to assessing contemporary virtue epistemology's theory of knowledge. He develops a novel theory of knowledge which, in his view, is more empirically adequate, better captures the ordinary conception of knowledge, and is ripe for cross-fertilization with cognitive science. He calls this view "abilism" and identifies several topics for future research.

Leading off Part VI, on understanding particular virtues, Michael Austin develops a Christian approach to humility. He claims that many analyses of humility, both religious and secular, emphasize its self-regarding aspects, including certain forms of self-knowledge, awareness of one's limitations, and an acknowledgment of dependence on others. Austin focuses on some underappreciated social aspects of Christian humility. He claims that on a Christian view, humility is important relative to how humans relate to God and one another. His primary aim is to offer an analysis of some of the social aspects of this virtue with a focus on relationships with other human persons. He also considers two objections to his account of Christian humility, in order to offer a limited defense of the account as well as a further explanation of its contents.

Angela Knobel, the leader of the theology side of the Character Project, focuses on developing a Christian approach to practical wisdom. She argues that Aquinas's account of Christian practical wisdom can accommodate Christians' own self-understanding in a way that more traditional Aristotelian accounts of practical wisdom do not. Because Aquinas incorporates a role for the Holy Spirit's guidance in the exercise of Christian practical wisdom, he is able to accommodate the Christian belief that even the virtuous are continually in need of God's guidance. Her chapter ends by addressing some challenges that must be met before contemporary accounts can accommodate Aquinas's insights.

Cristian Mihut's aim is to develop an account of the virtue of divine forgivingness grounded in a particular genealogy. Starting from the metaphor of *nasa awon* (bearing away transgression), Mihut argues that the Hebrew Bible sees God as paradigmatically a healer and forgiver of his people. He first unpacks the image of bearing away iniquity and gives four reasons for its centrality in the Hebrew Bible: its idiomatic frequency, paradigmatic frequency, liturgical depth, and fruitful conceptual overtones. After a methodological interlude, he develops an account of *nasa*-based divine forgivingness that involves the interlocking of two commitments: God's commitment to separate the offense from the transgressor and the commitment to absorb into God's self the consequences of the offense. Mihut also illustrates how this conception of divine forgivingness is at play in the story of Jonah, and ends by addressing two main objections.

Turning to one of the psychology projects, Angela Lee Duckworth and Eli Tsukayama propose a theoretic model that can explain both domain-general and domain-specific impulsive behavior, and discuss empirical studies that provide support for this model. They begin by discussing the relationship between self-control and character and provide some historical background on the domain-specific versus domain-general issue. The central part of their chapter presents some preliminary findings from their longitudinal investigation of domain-specific impulsive behavior. On their account, domain-specific impulsive behavior is a result of domain-specific temptation and, to a lesser extent, perceived harm. Their hope is that the findings reviewed in this chapter highlight the utility of a domain-specific approach in terms of improved prediction of life outcomes, understanding psychological processes, and targeted interventions.

Sara Konrath examines empathic character, understood as a set of interrelated dispositions, skills, motivations, emotions, and behaviors that involve a habitual responsiveness to others' needs. It is linked to higher prosocial behavior, lower aggression, and better health. There is much research demonstrating both its consistency within people over time and its malleability in response to environmental and situational cues, including face-to-face interventions. In her chapter, Konrath examines whether it is possible to increase empathic character using a mobile-based program. A large body of research in public health has used text messages to improve physical health outcomes and behaviors, but no research has examined whether text messages can be used to change traits. She conducted a study in which participants received six daily empathy-building text messages for a two-week period versus a control group. Konrath found that those in the empathy group showed some evidence of increased empathic character compared to control participants.

Part VII looks at how character develops and can be improved over time. The first three chapters examine character development in infancy, the elementary school years, and the college years, respectively. Jessica Sommerville investigates the development of moral concerns in the first two years of life, and specifically the development of distributive fairness concerns. She argues that infants' distributive fairness concerns emerge in the first year of life, and are related to infants' sharing behavior. Moreover, she demonstrates in her research that distributive fairness concerns in infancy share many features common to more mature representations of fairness: they contain evaluative components, vary in strength from individual to individual, and are weighed against competing social and moral concerns. Her findings provide evidence that at least some moral concerns are experientially influenced and identify potential mechanisms that support developmental changes and individual differences in fairness concerns, as well as point to ways to intervene to enhance positive character development.

Moving from infancy to the elementary school years, Elizabeth Boerger and Anthony Hoffman discuss how peer relationships are an important context for moral development, and yet children differ in the extent to which their behavior toward peers reflects concern with moral issues. They first provide an overview of individual differences in children's peer relationships including recent work on the social dynamics of bullying, and then discuss development in moral reasoning as it relates to children's understanding of the moral implications of prosocial and aggressive behavior toward peers. Against this background, they review the literature on two factors that may explain individual differences in children's behavior toward peers: theory of mind and moral identity. Finally, they examine how moral identity may interact with theory of mind to predispose children to either a predominantly prosocial or predominantly aggressive orientation toward peers, and discuss findings from their lab that provide initial support for the hypothesized interaction.

In Erik Noftle's work, early emerging adults aged 18 to 21 were tracked in a longitudinal study. Noftle investigated three aspects of character: character traits (Big Five facets), character strivings (strivings based on VIA character strengths), and moral self-attributes (self-evaluations of moral behavior). And he posed three questions: what is character like in early emerging adulthood and how does it change; how do aspects of character interrelate; and what do future projections of character—what he calls "character prospection"—reveal about which aspects of character are important, ideal, and expected to improve? According to Noftle, results indicate that early emerging adults see themselves as typically acting in good character but also as improving somewhat in college, and that those higher in moral self-attributes also have higher levels of character traits and strivings. Simultaneously, emerging adults see their ideal character levels as considerably higher than their current levels, expect character traits to improve in the near future, and view character striving as highly important.

Turning to religious approaches to character development, Tzvi Novick claims that insofar as Jewish normativity centers on law, it appears to have trouble making claims to shaping character. Put differently, it allows exemplars to play little or no apparent role. But traditional Jewish sources do make frequent use of exemplars in the domain of etiquette. According to Novick, etiquette combines, distinctively, two normative features not usually found together—namely, an intense interest in rule-following and a prominent role for exemplarity. In virtue of this distinctive coupling, he claims that etiquette plays an important role in enabling Jewish normativity both to remain law-centered and simultaneously to purport to shape character.

The next three chapters adopt a Christian approach to character development. Ray Yeo claims that character and virtue formation in the Christian tradition is distinct from its non-religious counterparts in its foundational insistence that genuine Christian character formation can only be achieved

through the gracious gift of God and cannot be had simply through human endeavors. However, he notes that the precise relationship between the transforming grace of God and the character transformation of the believer has not always been well understood. In light of this obscurity, Yeo attempts to provide a theoretical model for understanding the nature of the Spirit's indwelling in transforming grace by synthesizing Jonathan Edwards's theological psychology with William Alston's model of indwelling. The significance of the proposed model for Christian character formation is then explored.

In a closely related chapter, Elizabeth Agnew Cochran considers a Reformed Protestant theological approach which has historically affirmed that divine providence closely oversees particular individuals and events and that God alone is the author of salvation. This tradition's way of developing and articulating these claims poses challenges for contemporary Reformed ethicists working to consider the nature of moral agency in human beings. Her chapter proposes a character-centered model of moral agency that draws from the Roman Stoic understanding of moral responsibility but that simultaneously accommodates Reformed commitments regarding divine sovereignty and soteriology. On her view, an authentic conception of moral agency is embodied through the cultivation of a disposition of consent, an expression of faith or trust, to the workings of benevolent providence. In turn, she claims that this consent generates a second disposition through which agency is exercised: a disposition of love for all others in the universe, emerging from a recognition of our shared status as creatures of God.

Terence Cuneo's focus is on liturgy. Prominent within the ancient Christian liturgies, such as the Eastern Orthodox liturgies, is the activity of reenacting various episodes depicted in the Gospels, such as the burial of Jesus. He wonders how we should understand these reenactments and what they are for. His chapter is primarily devoted to answering the second of these questions, contending that a dominant aim of liturgical reenactment is to contribute to the construction of a narrative identity in the participants. Specifically, Cuneo's discussion is devoted to exploring the normative commitments that agents undertake when engaging in liturgical reenactment and how these commitments might contribute to the construction of narrative identities that, in turn, inform and sustain character traits of various kinds.

In the final chapter of Part VII, Elizabeth Bucar uses the concept of virtue to analyze the ethical significance and justification for the Islamic veil in the Indonesian context. She first provides an overview of Qur'anic directives to veil that connect modest dress to the cultivation of a good character, followed by a description of the particulars of fashion veiling in central Java with attention to how virtue is and is not invoked to justify certain styles. Finally, Bucar examines various themes that arise within a study of Indonesian veiling that might help ethicists to better understand the role of bodily practices, modesty, beauty, and politics in the cultivation of character.

Part VIII explores some of the serious challenges to virtue and character more generally which have been developed in recent years, especially challenges arising from neuroscience and from social psychology. For instance, Daniel McKaughan observes that virtue theories that appeal to character traits to explain and predict normatively evaluable behavior will eventually need to come to grips with the question of how such traits relate to a detailed biological understanding of their causes. Though many empirical questions relevant to such an integration remain unanswered, recent neurobiological accounts of the associations between genes, molecular-level biochemical pathways, complex social behaviors, emotions, and moral judgment have prompted claims that such work has profound implications for morality and our ordinary understanding of persons. McKaughan identifies ways that neurobiological accounts of the proximate causes of social behavior could, in principle, be relevant to virtue theory while resisting the hype that has surrounded recent research focusing on the neuropeptide oxytocin. He argues that so far, the results are open to interpretations that are not nearly as revisionary for our commonsense understanding of persons or rational action as some would have us believe.

The psychologist Matthias Mehl and his colleagues engage with the person-situation debate in social psychology and aim to shed light on the existence of moral character from a particular measurement perspective—namely, naturalistic observation. They examined the stability of moral daily behaviors using the Electronically Activated Recorder (EAR), a digital audio recorder that samples ambient sounds from participants' daily lives. Across three studies, participants wore an EAR for one or two weekends. Positive (e.g., empathy, affection, gratitude) and negative (e.g., bragging, blaming, arrogance) moral behaviors were coded from the audio files. In empirical analyses, the coded moral behaviors evidenced substantial temporal stability. Moreover, the stability of moral behavior was comparable to the stability of neutral linguistic behaviors (e.g., the use of articles and prepositions). Together, they take these findings to indicate that moral character can be naturalistically observed in daily life and is substantially stable across time.

David Gallardo-Pujol and his colleagues aim to better understand why people obey or disobey in strong and extreme situations. Using virtual reality simulations, they recreated Milgram's famous shock environment while measuring for, among other things, personality traits and intelligence. They found that low intelligence, high conscientiousness, and high aggressiveness are strong predictors of obedience in the Milgram setup. Their hope is that better understanding how these factors interact with situational forces can lead to a better understanding and prediction of human behavior in specific situations, especially when the potential for great harm is involved.

Further engaging with situationist themes, Matthew Talbert and Jessica Wolfendale note that some philosophers have recently argued that military

personnel typically are not responsible for war crimes. The reason is said to be that situational factors, such as battlefield stress and military training, undermine their capacity to recognize morally relevant features of their environment, and thus they should be excused for their wrongdoing. Talbert and Wolfendale challenge this conclusion and attempt to show that this account is inadequate as an explanation of war crimes and as a theory of responsibility. Drawing on social cognitivist accounts of personality, they describe how military training instills values, beliefs, and emotions in soldiers that inform their uses of violence in ways that can contribute to war crimes. They argue that even though military personnel cannot fully control the dispositions they develop through military training, perpetrators of war crimes are responsible for their behavior if their actions express morally objectionable attitudes that justify their victims in blaming them. Their account of responsibility thereby attempts to respect the victims' perspective in a way neglected by many competing accounts of responsibility.

Acknowledgments

This volume was made possible through the support of grants from the John Templeton Foundation and the Templeton World Charity Foundation. The opinions expressed in this volume are those of the authors and do not necessarily reflect the views of the Character Project, Wake Forest University, the John Templeton Foundation, or the Templeton World Charity Foundation. We would all like to gratefully acknowledge the support of the Templeton foundations, and especially the many contributions made by Michael Murray, John Churchill, Alex Arnold, Dan Martin, Kimon Sergeant, and Chris Stewart. Here at Wake Forest University we have been helped at numerous points with the funding competitions and eventual publication of this volume by Joshua Seachris, Kathleen Stimely, Lori Messer, Diane Samuel, and Amy Comer. We also thank the 12 reviewers of the full proposals from the funding competitions, who will remain anonymous but who have our sincere gratitude and appreciation.

We are very grateful to Peter Ohlin at Oxford University Press for his support of this volume, and Emily Sacharin for overseeing the publication process so helpfully.

Christian Miller would also like to thank his wife, Jessie Lee, and parents, Charles and Joyous Miller, for years of very wise support and advice. There is no way this project could have happened without their help. His department chairs, Ralph Kennedy and Win-Chiat Lee, his grant administrator, Joshua Seachris, and the philosophy department administrative assistant, Donna Simmons, were also great sources of wisdom and assistance. Finally, it has been a joy to work with Will, Mike, and Angela, and he is very grateful for all

the hard work and long hours they put into this volume and The Character Project in general.

R. Michael Furr and William Fleeson would also like to thank Kathleen Stimely for her years of dedication to The Character Project. They would like to thank the excellent postdocs Eranda Jayawickreme (now our colleague on the faculty at Wake Forest University), Erik Helzer, Laura Blackie, and Anna Hartley; graduate students Kelly Miskewicz, Alicia Jenkins, Pete Meindl, Ashley Hawkins, Max Barranti, David Mendrygal, Brandon Weiss, Raj Anderson, Cory Costello, Stephanie Komoski, Sarah Letson, Sarah Carper, and Michael Crowe; and the many wonderful undergraduate students— Kandace Andrews, Emma Wutschel, Evander Britt, Matthew Stanley, Anna Davis, Covington Paulsen, John Borkowski, Emily English, Scott Frankel, Anthony McEwen, Angelica Tesoriero, and Elayne Deelen—without whom our work could not have been done. We would also like to thank the Department of Psychology at Wake Forest for hosting The Character Project these many years, and Dale Dagenbach, Janice Jennings, Crystal Reid, Richard King, and Teresa Hill for their help with conducting the project. They would also like to thank Christian Miller for his many hours of hard work in making the project run. More personally, Will would like to thank his wife, Anne, and parents, June and Alan, for confidence and for the fascinating years of character-related conversations. Mike would like to thank his wife, Sarah, for her support and love.

Angela Knobel would like to thank her husband, Coy. She would also like to thank her dean, John McCarthy, and the staff at Catholic University's School of Philosophy for their efforts in support of this project. She would also like to thank the other participants in the project, especially Christian Miller, whose constant and conscientious work made this project possible.

PART } I

Overview of the Study of Character in Philosophy and Psychology

1 }

Some Foundational Questions in Philosophy about Character

Christian B. Miller and Angela Knobel

Research on character has exploded in recent decades, especially in the fields of psychology, philosophy, and theology. From discussions of virtue ethics in philosophy to the situationism debate in psychology to the role of the divine in moral formation in theology, interest in character seems to only be increasing with each passing year. And for good reason, since character and related notions like virtue and vice, are central for understanding so much of what we actually are like, as well as what kinds of people we should strive to become.

The chapters in this volume capture this spirit of energy and excitement. They are full of new ideas, innovative arguments, and empirical discoveries that will advance our knowledge of character on many fronts. Our goal in this first chapter is to provide a brief overview of what we consider to be some of the foundational questions in philosophy about character. We will not try to settle any of these questions but, rather, aim to provide some broad conceptual background on character that we hope will be useful in approaching the chapters which follow.

First Question: What Are Character Traits?

We can all come up with lots of examples of character traits. Heroism, cleverness, open-mindedness, cruelty, and compassion are just a few. As ordinarily understood, character traits are personality traits, and are concerned with the mental life of a creature—that is, the mental states and processes that constitute thinking.[1] These mental states and processes can and typically do

[1] It is an interesting and difficult question whether, on the other hand, all personality traits are character traits. I give some reasons for answering negatively in Miller 2014: chap. 1.

influence behavior in all kinds of ways. A shy person, for instance, *wants* to avoid speaking in public, and so may decline speaking invitations because of that. A sociable person, on the other hand, may get *excited* by an upcoming party and will spend hours mingling with the crowd.

It turns out that it is hard to say much more about what character traits are without first taking a stand on a debate between three rival positions: the dispositional view, the summary view, and the conditional view. The first of these positions is by far the most popular approach in philosophy and theology, but it is more controversial in the psychology literature. We will focus on the dispositional view in this chapter, given its overall prominence.

On the dispositional view, character traits are more than just momentary states of mind; and so, just because someone is exhibiting certain thoughts, it does not mean that he has the relevant underlying trait. Character traits can give rise to characteristic mental *states*, but mental states do not have to depend on character traits.[2] As an illustration, Smith might be in a compassionate frame of mind—he might be thinking about how best to help someone in need, care a lot at this moment about helping that person, and arrive at the correct answer as to what would be best to do for her. It certainly seems that Smith has compassionate mental states. But if this happens to be the only time in his life when he thinks and feels this way, and if otherwise he just tries to promote his own self-interest, then we suspect we would also likely not say anything more than this—he does not have a trait of compassion, but just a momentary compassionate state.

The same idea applies to the distinction between character traits and characteristic bodily *actions*. Smith might make a large donation to charity. In most cases we would call that a compassionate action, and would praise him for it. But merely knowing this about his action does not necessarily tell us anything about either the mental states behind the action or the traits (if any) which led to it being performed. For all we know, Smith could have been entirely in a selfish frame of mind, wanting to be recognized by society. That state of mind could have arisen from a trait of selfishness. So, a compassionate action does not entail the possession of a compassionate trait. And neither does a compassionate state of mind.[3]

---

[2] For the distinction between traits and states in psychology, see, e.g., Mischel and Shoda 1995: 257; Fleeson and Noftle 2008a: 1358; 2008b; Fleeson and Gallagher 2009: 1099; and Roberts 2009: 140 (and the references cited therein). The distinction is commonplace in philosophical discussions of character.

[3] Here we claim, somewhat controversially, that a person could perform a compassionate action from either a good or bad motive—trait properties of actions are not tied to the motives or intentions behind the actions, on our view, just as they are not tied to any character traits which give rise to them either. Rather, we prefer to say that performing a virtuous action just amounts to the person's performing that action, whatever it happens to be, which is deemed to be a virtuous action by the correct normative theory, regardless of what motives the person had for performing it. For instance, it might be the action that has the property of being what a fully compassionate person, acting in character, would have also performed in the same circumstances, as Aristotle seems to hold (*Ethics* 1105b6). But again, we recognize that this is controversial, and nothing of substance hangs on it in what follows.

Thus, for someone like Tom to have the character trait of shyness, for instance, he has to have some enduring *tendency* or *disposition* to have shy thoughts and act in shy ways. This disposition is distinct from the shy thoughts and actions, although it can give rise to both. Furthermore, these thoughts need not be active all the time or in every situation; when he is alone, for instance, Tom may not have them at all. Rather, it might only be when he is in certain conditions which are relevant to this trait, such as parties or classrooms, that they kick in and play an active role in his psychology. As philosophers like to say, in these conditions Tom's shy thoughts go from being merely *dispositional* thoughts, to being *occurrent* thoughts.[4]

Similarly, the shy beliefs and desires that Tom forms in these situations need not be precisely the same *particular* mental states on every relevant occasion—they can be as diverse as wanting to leave a party or wanting to hide behind a large football player in class. What matters is that they are of the same broad type, namely, shy thoughts, not that they are mental states with exactly the same content on each occasion.

So, generalizing from the example of Tom, the predominant way in philosophy and theology of understanding what it is for a person to have a character trait is, roughly:

> (1) A disposition to form beliefs and/or desires of a certain sort and (in many cases) to act in a certain way, when in conditions relevant to that disposition.[5]

A person who is shy is disposed to believe, desire, and act shyly, and can form such thoughts and act this way when, for instance, at a large party. Someone

---

[4] Occurrent thoughts, though, need not be conscious. I can have many subconscious occurrent desires which causally influence my behavior in all kinds of ways. For instance, Tom might be influenced to leave the party by a desire to avoid crowds of people, without realizing that he in fact has this desire.

[5] Talk of 'beliefs' and 'desires' is intended very broadly, and so would include emotions and intentions as well. Also, the ambiguity about whether character traits pertain to beliefs, desires, *and* actions is intentional since, while some character traits certainly do seem to involve all of these elements, as in the case of compassion or selfishness, others may not. Possible examples of a trait without a belief element might include irritability or general anxiousness, while foresight or closed-mindedness might not have a desire element.

Finally, note that nothing in (1) requires that there be individual differences between how people are disposed to believe, desire, and act in order for there to be personality traits. Psychologists often focus on individual differences when studying traits, but we claim this focus should be restricted to *gathering evidence* for their existence, rather than as a conceptual requirement for understanding them in the first place. In principle, there may be no individual differences in a population's possession of a personality trait—it is conceivable that everyone could have honesty or courage equally, for instance. Hence individual differences are not constitutive of such traits in the first place. For a similar claim, see Doris 2002: 19n23; Kamtekar 2004: 468; Badhwar 2009: 280; and Sosa 2009: 287. For the opposing view, see Johnson 1997: 74, 87; Funder 2008: 570; and Fleeson and Noftle 2009: 151.

who is sociable is typically disposed instead to form different thoughts and exhibit different behaviors when at large parties.[6]

A few additional points about these dispositions. It is a feature of dispositions in general that they are sensitive to certain *stimulus events* or *stimulus conditions* specific to the given disposition. In virtue of being fragile, for instance, a vase might be sensitive to being hit by a baseball, but not to the color of the baseball. Similarly, a properly functioning thermostat is sensitive to the temperature in the room, but not to its smell. It seems that certain events and facts about a situation or environment are relevant to a disposition in a way that others are not.

Stimulus events can trigger characteristic *manifestations* of dispositions. A baseball can break a fragile vase, and the temperature of the room can lead to a certain reading on the thermostat. Prior to the stimulus, the disposition might have been latent, as, for example, the disposition to believe that $2+7=9$ is latent during most of the day. But with the right kind of stimulus, such as a math test, the disposition can become manifest, say in the form of the occurrent belief that $2+7=9$. Indeed, all dispositions to believe or desire certain things become manifest directly in the form of occurrent mental states. These occurrent mental states can then lead, in many cases, to corresponding behavior.[7]

On this approach, character traits work in the same way. The trait of compassion can be triggered by what the person sees as ongoing suffering, and can lead directly to compassionate thoughts and ultimately to compassionate behavior aimed at relieving that suffering.[8] Note that the immediate stimuli for character traits do not have to be the actual features of a situation, but can just be the person's *impressions of* those features, impressions which in some

[6] For representative examples of the link between character traits and dispositions in the philosophical work on character, see the references cited in Miller 2013: 7n12; and Doris 1998: 509n20; 2002: 15n2.

[7] Things are not so simple, of course. Background conditions also play an important role. A match has the disposition of being flammable and so typically lights (manifestation) when struck (stimulus), but not in an environment without oxygen. A vase has the disposition of being fragile and so typically breaks (manifestation) when knocked off a ledge (stimulus), but not if gravity is much weaker than it is on Earth. And so forth.

Note that these are cases where the background conditions interfere with the ability of the disposition to be manifest, but still allow the disposition to continue to exist. Other background conditions may either prevent or eliminate a disposition from being instantiated in the first place, such as a healthy human being with the disposition of being strong suddenly passing away from a heart attack. See Mumford 1998: 86.

[8] But not if there are interfering background conditions, i.e., the person is experiencing serious depression or a mental illness which blocks the operation of the trait and precludes the formation of compassionate thoughts even in the presence of obvious suffering. For general remarks about character traits and background conditions, see Brandt 1970: 35; and Upton 2005: 135–136; 2009: 27–30. In addition, nothing about the foregoing is meant to imply that the compassionate behavior need depend *only* on the trait of compassion—other character traits, for instance, can also play a partial causal role as well.

cases can be seriously mistaken and yet still activate the relevant trait. Suffering that goes unnoticed in a situation through no fault of the person's own, for instance, will not have a bearing on her trait of compassion, even if it is true that if the suffering were noticed, then she would typically help to relieve it.[9]

In functioning in these ways, character dispositions are not inert properties in a person's mental life; rather they *causally mediate* between their various stimuli and manifestation events. So if Jones has a compassionate disposition, whereas Frank does not, that is not merely an individual difference between them. It is also (and crucially) a difference in the possession of a property which (together with the relevant stimulus) plays a significant causal role in leading to Jones's compassionate thoughts and behavior.

Because of this causal role, trait dispositions can also create expectations about how a person would likely behave in various situations, thereby serving as a basis for making accurate predictions. If, for instance, George Washington was deeply courageous and Abraham Lincoln highly honest, then we can reasonably predict how they would likely have behaved in different circumstances during the course of their lives. Similarly, for compassionate Jones, the following might be true:

(2) If Jones were to encounter a person who he notices is in need of a moderate amount of help, he would typically attempt to help, regardless of whether helping is in his self-interest.

Whereas, for selfish Frank:

(3) If Frank were to encounter a person who he notices is in need of a moderate amount help, he would typically attempt to ignore the need, unless he thought helping was in his self-interest to a relevant degree.

Of course, as stated, these predictions are much too simplistic. Background conditions can interfere with dispositions in all kinds of ways—Jones might also be suffering from depression or mental illness, for instance. But the important point here is that the causal activities of trait dispositions can create expectations about the person's future behavior in both himself and others.[10]

---

[9] For this last point, see Alston 1975: 21. Note that nothing about the foregoing implies how *frequently* character traits are actually triggered—that will depend on a number of factors, including how strongly the trait is held, the situations the person is in, and even the nature of the particular trait itself. Some traits, such as fairness, perhaps, or modesty, are by their very nature likely going to be highly active during the day for people who have them to a significant extent. With other traits, such as perhaps bravery or integrity, the person may rarely be in the kinds of situations which trigger their activation. Indeed, it seems possible on conceptual grounds that a person could have a character trait which is never triggered at all during her lifetime.

[10] Various attempts have been made in the literature to improve these predictions. For an overview, see Upton 2009: chapter 2.

To summarize, dispositions in general can be stimulated in various ways, and so long as there are no interfering background conditions, they can give rise to their relevant manifestations. If Jones has a particular trait disposition, such as compassion, then Jones has an actually existing (instantiation of a) property which can causally bring about occurrent compassionate thoughts and, thereby, compassionate behavior. These dispositional properties ground the truth of conditionals about what a person would likely think and do in relevant situations.

As noted, the idea that character traits are dispositions should be an acceptable claim for most philosophers. But among psychologists, two opposing positions have emerged, each with its own supporters.

According to what can be called the 'summary view', trait terms refer only to *actual patterns of* relevant mental thoughts and bodily action. A person's character traits on this view, then, just are the relevant mental states and bodily actions that she has in fact exhibited over time. There are no dispositions over and above these mental states and actions which might have been responsible for bringing them about in the first place. For instance, David Buss and Kenneth Craik maintain that "the statement 'Mary is arrogant' means that, over a period of observation, she has displayed a high frequency of arrogant acts, relative to a norm for that category of acts."[11] Similarly, John Johnson claims that "[t]raits are consistent patterns of thoughts, feelings, or actions that distinguish people from one another."[12] Walter Mischel states in his 1973 paper that "the present position sees [traits] as the *summary labels* (labels, codes, organizing constructs) applied to observed behavior."[13] And Jerry Wiggins nicely summarizes the view by remarking: "[W]hen we say that John is aggressive, we are asserting that the general trend or disposition of his conduct, to date, has been to engage in a variety of aggressive actions over a period of time."[14] Note that even though Wiggins mentions the term 'disposition', all that term refers to on this view is the aggressive behavior the person has already displayed up to this point in time.

Advocates of the rival 'conditional view' agree with the summary view that there are no traits understood as properties of human psychology which play either a causal or an explanatory role in relation to thought and action. But they nevertheless maintain that these traits can still be *ascribed*

[11] Buss and Craik 1983: 106.

[12] Johnson 1997: 74.

[13] Mischel 1973: 264 (emphasis his). To be fair, this quote does not say that this is *all there is to* traits, although that seems to be the natural implication of Mischel's remark and also what is suggested by the surrounding context.

[14] Wiggins 1997: 108. For helpful and sustained defenses of the summary view, see Buss and Craik 1983; and Wiggins 1997: 102–108. For other remarks which seem to be in line with the summary view, see also Hampshire 1953: 9; Mischel 1968: 42, 68–69; 1973: 262, 264; Hogan et al. 1977: 256–257; Buss 1989: 1383, 1386; Hogan 1991: 875; 1996: 170–173; Revelle 1995: 315; Winter et al. 1998: 232–236; Fleeson 2001; 2007: 825; Fleeson and Noftle 2008a: 1363; 2008b; and Heller et al. 2009: 171.

to individuals in virtue of a person's potential behavior. In other words, this view is centered on *counterfactual* rather than *actual* behavior—that is, how Jones *would* think and act if he were to notice a moderate need for help, not how he *has* thought and acted in the past. Thus this position is an alternative to the summary view.[15]

As we indicated at the start, we will refrain here from weighing in on this and other lively disputes in the literature. Of course, an expression such as 'personality traits' or 'character traits' can be defined as one chooses, and even in ordinary thought there are likely multiple senses of these phrases. Nevertheless, there may be important costs and benefits to employing one of these frameworks as opposed to the others. For the remainder of this introduction, we assume the dispositional approach for convenience; but everything that follows can be reformulated in terms of the other two approaches.

Second Question: What Makes Up a Character Trait?

According to the dispositional view, character traits are causal dispositions which can exist in our mind. But what are these dispositions like? And how are they related to the rest of our mind? Fortunately, here there is broad consensus. Character trait dispositions are, at least to a large extent, made up of interrelated mental state dispositions.[16] Here, we try to explain this idea.

We have said that a trait like compassion is itself a causal disposition. But we can also have very specific mental state dispositions as well that are relevant to compassionate behavior. Some of these might include:

A disposition to form the belief that it is important to help a person in need.

---

[15] Despite the non-existence of trait dispositions, trait *ascriptions* on this view are still true or false such that, for instance, it is true of Jones that he is compassionate, and it is false of him that he is selfish. The familiar strategy here is to reduce ascriptions of trait dispositions to conditional behavior (whether mental or physical) and hence to conditionals, such as the following overly crude example:

> (CV) The ascription 'Jones is compassionate' is true if and only if (and because) if Jones were to encounter one or more persons whom he notices is in need of a moderate amount of help, he would typically attempt to help and do so from a compassionate state of mind.

Note that nothing about this statement mentions a property of compassion—the ascription can be understood entirely, not in terms of properties of Jones's mind, but simply in terms of his counterfactual behavior. Thus, by providing some suitably sophisticated conditional analysis of ascriptions of trait dispositions, the conditional view hopes to avoid any commitment to actual trait dispositional properties themselves. For relevant discussion of the conditional view of traits, see Wright and Mischel 1987: 1161–1162; and Newman and Uleman 1989: 163–164.

[16] For references to the numerous philosophers who have advanced this idea, see Miller 2014: 26n63. The qualification "at least to a large extent" is important, as certain abilities, faculties (e.g., willpower if it is a faculty), or skills (e.g., perceptual sensitivities) might also be needed in at least some cases for the possession of a character trait, and perhaps these abilities, faculties, or skills do not themselves entirely consist of mental state dispositions.

A disposition to form the desire to help another person for his or her own sake.

A disposition to form beliefs about what are and are not appropriate forms of helping.

The question, then, is this—how do these specific belief and desire dispositions relate to the following?

The disposition of compassion.

One answer could be that they are completely unrelated, so that whether a person has the trait of compassion has nothing to do with whether she has dispositions to form compassionate beliefs and desires. But no one has ever held such a position as far as we are aware, and for good reason.[17] Rather, the standard answer is to say that, in general:

> A character trait disposition which is had by a person *consists of* some cluster of her relevant interrelated[18] mental state dispositions such that necessarily, if she has this cluster of dispositions, then she instantiates that character trait as well in virtue of having those dispositions.[19]

In other words, having the relevant mental state dispositions *enables* a person to have the corresponding character trait. The trait, to put it differently, is anchored in the specific dispositions to form relevant beliefs and desires.[20]

Returning to Jones, his trait of compassion is itself grounded in underlying mental state dispositions in Jones's mind, and, specifically, in mental state dispositions which are *appropriate* to the virtue of compassion from an ethical perspective. These could include his disposition to *recognize* people suffering in his environment, to *want* to help relieve their suffering regardless of whether doing so would benefit himself, to *weigh* different helping strategies, and so forth. But it would *not* include, for instance, a disposition to want to make people feel guilty after being helped, since from any plausible ethical perspective that would not be a disposition which is an appropriate constituent of the

---

[17] One reason, presumably, is simply the intuitive implausibility of thinking that whether a person has the trait of compassion is independent of what her motives for helping others are like, and so could be compatible even with having only the crudest possible desires for her own self-advancement.

[18] That the mental state dispositions are 'interrelated' is important. Obviously, not just any random collection of mental state dispositions will amount to a character trait like compassion independently of the content of those mental states. But it is also not enough for a person to just *have* the appropriate mental state dispositions. A person does not have the trait of compassion unless she has the relevant belief and desire dispositions, *and* they (or the occurrent states they give rise to) trigger each other when appropriate. The beliefs that someone is in need and that I can help that person should, in relevant cases, be related to a desire to devise means of helping him. Otherwise, there could be an unrelated bunch of mental states, rather than a character trait. For similar remarks, see Snow 2010: 20.

[19] For an important qualification of this claim, see footnote 16 herein.

[20] What, precisely, is this anchoring relation? Is it the relation of identity, or perhaps a non-reductive constitution relation? Here, we remain neutral on this lively debate between trait monists and trait dualists, but see Miller 2014: chap. 1.

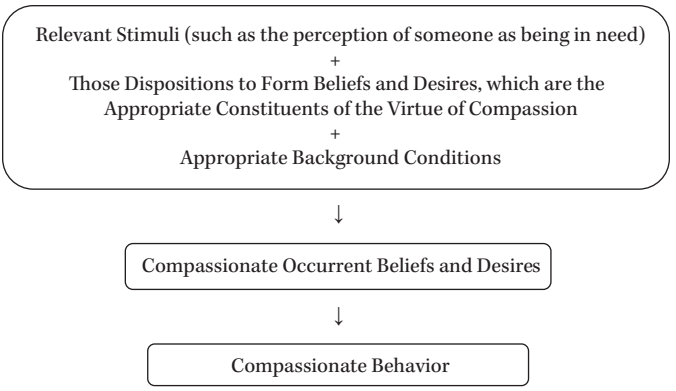
Relevant Stimuli (such as the perception of someone as being in need)
+
Those Dispositions to Form Beliefs and Desires, which are the
Appropriate Constituents of the Virtue of Compassion
+
Appropriate Background Conditions
↓
Compassionate Occurrent Beliefs and Desires
↓
Compassionate Behavior

FIGURE 1.1 Compassion, Compassionate Thoughts, and Compassionate Behavior.

virtue of compassion. Figure 1.1 provides an illustration of the picture we have in mind; the arrows symbolize causal influence.

So when Jones encounters someone in need and the relevant background conditions are cooperative, his trait can be activated. This amounts to the underlying dispositions to form mental states of various kinds that are appropriate to the virtue of compassion being activated. As dispositions, they will have their own stimulus conditions. And, as the grounds of compassion, they can be activated by the perception of need. In being activated, they can go from being latent dispositions to causing the formation of occurent mental states as their manifestations—beliefs, wishes, hopes, intentions, and so on, of the relevant sort, in this case, compassionate—which in turn can cause the performance of compassionate actions, other things being equal.

Third Question: What Are Some Additional Features of Character Traits?

Because they show up in a number of chapters in this volume, we want to focus on four additional features of traits: minimal thresholds, degrees, stability, and cross-situational consistency.

In order to possess a character trait, a number of necessary conditions have to be met. For instance, the person must first possess one or more psychological dispositions. This is one of the necessary conditions for being eligible to have a character trait of *any* kind in the first place. But there are additional necessary conditions that must be met in order to have a *specific* character trait, say, of compassion rather than cruelty. Call these specific conditions the *minimal threshold* a character trait has to meet in order to qualify as that particular kind of trait rather than some other.

This idea should be familiar from the previous discussion. If character traits are grounded in specific mental state dispositions, then one cannot count as having a given character trait, such as honesty, without first having the specific dispositions to form beliefs and desires relevant to honesty, and also not having the specific dispositions to form beliefs and desires relevant to dishonesty. So a given character trait qualifies as a given virtue provided that its underlying mental state dispositions are of that virtuous kind themselves.

To make this more concrete, consider compassion once again. If a politician reliably helps in a variety of helping-relevant circumstances, then it might seem as if he has the trait of compassion. But suppose it turns out that he desires to help others only when it makes him look better to the media. Then, while this trait might satisfy one of the necessary conditions in the minimal threshold for compassion (it leads to helping behavior), it has not satisfied another one—the motivation is not of the appropriate sort. Note that the claim here is *not* that the trait therefore only qualifies as being weak compassion as opposed to deep compassion. Rather, the claim is that the trait does not (yet) qualify as compassion *at all*. Until his trait satisfies all the necessary conditions for *compassion*, it does not meet the minimal threshold for compassion, and so the person does not instantiate the specific trait of compassion to any degree at all.

To use another example, for some character traits there are certain obvious situations in which, other things being equal, one would expect a person with the trait to exhibit trait-appropriate behavior, even if the trait is held to only a weak degree. As Richard Brandt notes for the trait of sympathy, for instance, "Each of us has a conception of what a sympathetic person will do in a situation of a certain kind. For instance, we may think a sympathetic person will interrupt a friendly tennis game to tend to a child who has fallen off his bicycle. If he won't do that, we do not call him sympathetic."[21] So reliable performance of sympathetic actions in these obvious cases seems necessary, again other things being equal, for even qualifying as a person with the trait of sympathy.

Now if a person does not have a trait like sympathy or compassion, it does not follow that he must have acquired an *opposing* trait, such as selfishness or cruelty, to any degree. She may simply have none of these traits at all. Furthermore, we ordinarily think that character traits are not permanent features of our personalities. We assume, in other words, that we can gain and lose them, at least over an extended period of time, and so while a person might have had the trait of sympathy ten years ago, perhaps he has lost it in the interim.

In our experience, the idea of a minimal threshold for possessing a given character trait is relatively uncontroversial to philosophers and theologians working in ethics, but is less plausible to many personality and social

[21] Brandt 1970: 36. For brief criticism of this proposal, see Butler 1988: 237.

psychologists.[22] To use terms common from psychology, our claim is that character traits are both *categorical traits*—you have to meet certain standards to qualify as possessing them—and (once you do qualify) *continuous traits*, which can show up anywhere along a wide continuum of more and less. Psychologists, we often find, sometimes grant the distinction between categorical and continuous traits, but deny that character traits are categorical ones.

Despite this disagreement, all parties seem to agree that particular character traits can come in *degrees* of more or less (i.e., they are continuous traits). Two people can have the trait of compassion, say, but they might each have it to a different degree.[23] So in light of this discussion, there are two distinct questions that one can ask about a person's character: does she have a particular trait, such as honesty, *at all*, and if she does, *to what degree* does she possess this trait?

Let's turn to the third feature of character traits, namely, *stability*. Suppose Jones's compassion one day leads him to donate money to the Salvation Army representative on the street corner. And he does this again the next time he comes across this person. And again, the next time. Jones is thereby exhibiting stability of a certain kind, which is what can be called:

> *Single-Situation Trait Stability*. A person regularly manifests a certain type of behavior which is in accordance with the trait in question, over several instances of the same trait-relevant situation. So, for instance, if Jones is deeply compassionate, he might donate money to the Salvation Army representative every time he goes to the grocery store this month.

Now consider another form of stability. Instead of focusing on just one situation, consider Jones's behavior over a variety of situations and over the course of a given time period, say a week or a month. And then compare this *aggregate* or collection of specific behaviors, to the aggregate of behaviors related to the same trait in one or more additional periods of time, say the next week or month. If the average level of his aggregate trait behavior is similar in these two time periods, then he is exhibiting a more general form of stability:

> *Aggregate Trait Stability*. A person regularly manifests similar *levels* of aggregate behavior which is in accordance with the trait in question, over different periods of time. So, for instance, while Jones might have helped in different ways during two different time periods, his average level of compassionate behavior across all these helping situations might still be very similar.

---

[22] For discussion that is broadly related to the notion of a minimal threshold, see Hursthouse 1999: 149–150; Swanton 2003: 25; Upton 2009: 7, 49, 53; Kupperman 2009: 245; Eklund 2011; and, especially, Brandt 1970: 33–37; and Russell 2009: 112–117. As Russell rightly points out (p. 114), there will likely be some degree of vagueness rather than a sharp boundary between whether a trait just barely qualifies as a particular kind of trait and whether it does not qualify.

[23] This assumption about character traits is widespread. For references see Miller 2013: 15n26. The degree to which a person has a character trait is presumably some complex function of the degree to which he has the underlying mental state dispositions and the relations between them.

In addition to multiple kinds of stability, there are multiple kinds of *consistency*.[24] Return to compassionate Jones and his donating money. That is not the only form of compassionate behavior most people would expect him to exhibit on any given day. Generally, and other things being equal, in a situation involving a severe wreck on a deserted road, he would be expected to call 911; in a situation involving dropped papers, he would be expected to offer to help pick them up, and so forth. So Jones would be expected to be cross-situationally consistent throughout the day in his compassionate behavior:

> *Momentary Cross-Situational Consistency*. A person manifests behavior which is in accordance with the trait in question, while in several different trait-relevant situations during a narrow period of time. So, for instance, Jones might offer to pick up dropped papers, help make change for a dollar, and hold the door for a handicapped person, all in the span of an hour.

Of course, we can wonder not only about how cross-situationally consistent Jones is in his behavior during an hour or a day, but also during longer periods of time:

> *Extended Cross-Situational Consistency*. A person regularly manifests behaviors which is in accordance with the trait in question, while in a variety of different trait-relevant situations during an extended period of time. So, for instance, Jones might perform compassionate actions in many of the situations he encounters during the span of a month.[25]

It is worth stressing that for there to be cross-situational consistency, the actions do not need to be precisely the same ones in the different situations (such as giving five dollars no matter what the helping-relevant circumstances are, which would be absurd), but rather, need only be different instances of one broad type of character-based action (such as compassionate action). In other words, the trait of compassion can give rise to such diverse particular actions as donating to charity, picking up dropped papers, and taking someone to the hospital.

---

[24] On some taxonomies, 'consistency' is the very broad label, one type of which is stability. Furthermore, it is important to stress that there are many other kinds of consistency/stability which are also important for discussions of character beyond those mentioned here. For instance, Fleeson and Noftle (2008a) distinguish between 36 kinds of consistency which pertain to personality traits in general, and go on to argue that no one kind is necessarily privileged. At the same time, the four kinds of consistency/stability on which we focus here, have historically been, and indeed deserve to be, prominent in both conceptual and empirical discussions of character traits.

[25] This is different from aggregate trait stability, which only concerns the similarity of the person's behavior over time, and not whether that behavior is consistent from one situation to the next. Jones might be very stable in the average level of his helping behavior from one month to the next, but only help in certain situations and not others.

With these four kinds of consistency in mind, it is a mistake to think it essential to character traits that they be both stable *and* cross-situationally consistent. *Local* character traits are ones which require stability in relevant behavior with respect to one kind of situation, but fail to exhibit cross-situational consistency. Conrad, for instance, unlike Jones, might repeatedly make a donation to charity when asked, while exhibiting little else in the way of compassionate behavior. He might, in other words, have what can be called 'donation compassion', whereas Jones has the *global* trait of compassion in general.[26] So cross-situational consistency is not a necessary feature for qualifying as a character trait.[27]

Finally, we can end this section with a very important and widely used distinction for thinking about consistency. When a person is said to be consistent in his compassionate behavior in the same situation on two different occasions, what counts as the 'same' situation? Is the 'same situation' defined by that person himself, for instance, or by an outside observer? Similarly, when a person is said to be consistent in his compassionate behavior across two different situations, who decides that they count as 'different' situations?

As indicated, one approach to answering these questions is to use a perspective that is external to the person in question, perhaps involving the beliefs of ordinary independent observers, or, in the case of psychology experiments, the perspective of the psychologists who are conducting the experiments. Such observers might consider a person's home and workplace, or dorm and classroom, to be two different types of situations, and then go on to test whether her behavior in question is consistent between them. Thus, if she ends up failing to act compassionately in both kinds of situations, then she does not exhibit (momentary) cross-situational consistency.

Using the standard terminology in parts of the psychology literature, such an approach appeals to the *nominal* features of the situation. But another way of thinking about consistency relies on what features of situations are important to the person in question. These are the *psychologically salient* features, or "the features of the situation that have significant meaning for a given individual."[28] Note that psychologists do not limit these features to those which the person thinks or consciously takes to be salient.[29] Rather, these are features which cause relevant psychological states to activate in the person's mind, whether she is conscious of their impact or not. Thus racial features, for instance, could

---

[26] For more, see Doris 1998: 507–508; 2002: 23–25, 64–66. Actually, this is only one way of construing 'local' traits. They could instead be understood, not as local to a kind of situation, but as local to a period of time or to some other criterion. We have chosen the approach described here since it is the most common in the philosophy literature on character. Thanks to William Fleeson for this reminder.

[27] For this point in psychology, see Johnson 1997: 75; 1999: 444, 446; Mischel and Shoda 1998: 235; and Mischel 1999: 457. For an opposing view, see Buss 1989: 1385.

[28] Mischel 2004: 15. The distinction between nominal and psychologically salient features of situations is especially prominent in social-cognitive approaches to personality. For extensive discussion and citations to the literature, see Miller 2014: chapter 5.

[29] Thanks to Rachana Kamtekar for this reminder.

trigger implicit stereotypes without the person even realizing what is going on; nevertheless, these features would still count as 'psychologically salient' in the sense that is relevant to most researchers in this area.

These two ways of thinking about consistency can lead to significant differences. Two situations might seem to have nothing to do with each other, but for the person in question they might both have similar psychologically salient features which end up leading to consistent behavior. Or, in the opposite manner, two situations might, according to the psychologists conducting a study, seem to be quite similar in their relevance to a given trait, but end up having very different psychologically relevant features for the participant who experiences them. This can lead the experimenters to conclude that the participant does not exhibit cross-situational consistency in his behavior in the area, even when he is indeed being highly consistent relative to his own mental life. Bem and Allen offer a nice example of how this might happen in an actual study: "[A]n individual who regards himself as extremely conscientious might not consider his casual attitude toward personal hygiene as pertinent to that trait. The fact that the investigator's concept of conscientiousness might include personal hygiene within it is not relevant."[30]

Fourth Question: Why Do We Appeal to Character Traits?

Thus far we have seen some ideas about what character traits are, how they are constituted, and what some of their causal properties are. But none of this speaks to their importance. Why, in other words, do we tend to invoke character traits in our thought and speech? Naturally, there are many reasons, and here we want to highlight five of the central ones:

(a) *Understanding*. Character traits are a basis for understanding ourselves and others by classifying people in various ways which can be important to interpersonal and intrapersonal interaction. When I form an impression of Smith, for instance, my general understanding of him may be framed in terms of character traits and shaped heavily by my perceptions of his honesty, compassion, or generosity. The same could be true, not only about Smith, but about myself as well.[31]

(b) *Explanation*. Character traits are a basis for (partially) explaining why people act the way that they do. Smith may cheat on tests because he is dishonest (either in general or just dishonest in test-taking

[30] Bem and Allen 1974: 511. As Eaton et al. note, "[T]he situations that interact with personality dispositions are not relegated only to the externalities with which one is in contact. Instead, features of situations can be *created* by an individual's thoughts, mood states, fantasies, and so on." (2009: 212; emphasis theirs).

[31] Thanks to R. Michael Furr for pointing out this role. See also Hogan 1996: 170.

situations), whereas Jones may regularly give money to charity because he is compassionate, other things being equal.[32]

(c) *Prediction*. Character traits are a basis for predicting what a person who has them will likely do in the future.[33] For instance, if I think that Smith is shy, then I can take myself to have a fairly good idea of how he will likely behave in a room full of people at the next work party, other things being equal.

(d) *Evaluation*. Character traits are a basis for normatively assessing a person. So when I say that Smith is dishonest, I am evaluating him as a person in a negative way, at least in that one respect.

(e) *Imitation*. Character traits are a basis upon which to imitate another person and cultivate positive traits while avoiding negative ones. So I might strive to become more compassionate, and model my life after such people as Jesus, Gandhi, Socrates, and Mother Theresa, who are said to have had certain virtuous character traits to a high degree.[34]

In addition, among philosophers and theologians, character traits are often invoked in constructing a normative ethical theory. How central a role traits play will vary from theory to theory. In some, they are more peripheral, such as versions of utilitarianism which mention traits only insofar as they are important for disposing people to maximize utility. In other theories, they are the central notions, such as versions of Aristotelian virtue ethics which ground obligatory or right actions in the behavior of a person with virtuous character traits.

Fifth Question: What Are Some Normative Differences between Character Traits?

At a very general level, we know that there is a distinction between the normative and the descriptive. Here are some descriptive statements:

"$2 + 2 = 4$"

"A water molecule is composed of two hydrogen atoms and one oxygen atom."

"George Washington was the first president of the United States of America."

---

[32] This explanatory role is not limited to specific behaviors. Richard Brandt notes that historians "make free use of trait-concepts in explanations of events (the collapse of a nation may be ascribed to things like the queen's vanity)" (1970: 23).

[33] But only what a person is 'likely' to do, since among other things people seem to be able to act out of character as well. In those cases, explanations in terms of character traits would also not be helpful.

[34] Note that the functional roles in (a) through (e) are roles which character traits play, regardless of whether anyone in fact thinks about them in these ways (or under these descriptions). In other words, a character trait serves as a basis to explain and predict someone's behavior, even if no one realizes that the person has that trait. For further discussion of one or more of the above functional roles, see the references cited in Miller 2013: 13n24.

"Benjamin Franklin was the first president of the United States of America."

These statements, as descriptive, only aim to capture the way the world actually is. Not all of them succeed in doing so (such as the fourth), but the key concern is whether the world is as each statement says it is.

Normative statements, on the other hand, are concerned with evaluating the way the world is or how the world ought to be:

"There should be more charity organizations."

"He is a cowardly person."

"The criminal deserves the death penalty."

"If only he had acted for better reasons."

Sometimes the descriptive and the normative come together. The statement that a person is cowardly may at once seem to describe the person's actual character and also serve to evaluate him as a person. But we know that the descriptive and normative domains are distinct. For, as a matter of fact, the world as it is falls short in all kinds of ways from the way that it should be. No examples are needed for such an obvious point.

As noted with the previous question, character traits are a basis for normative evaluation, and not just of a person but also of his or her occurrent thoughts and actions. It is tempting to focus on *moral* evaluation in this context, but there are in fact a variety of different normative domains with character traits pertaining to each of them (and in some cases, applying to more than one domain). Here is a fairly standard list of some of the normative domains and their relevant traits:

Moral character traits, such as honesty, compassion, and temperance.

Epistemic character traits, such as intellectual humility and intellectual courage.

Prudential character traits, such as diligence and cleverness.

Religious character traits, such as religious faith and religious devotion.

Athletic character traits, such as competitiveness and discipline.

Other categories could be added, too, such as aesthetic character traits. And there may be controversy over particular categories, such as whether the religious and the moral should be distinguished (on some religious frameworks, they would not be).

Fifth Question: Do Character Traits Actually Exist in Human Beings?

This question is intended to be understood psychologically, as the question of whether human beings have such traits as part of their psychological lives. It is a question best pursued on empirical grounds, and as such falls more under the purview of psychologists than philosophers and theologians.

And it is a question which is easy to answer. Aside from extreme versions of the situationist movement in psychology in the 1960s, the overwhelming consensus has been yes. Even leading figures associated with situationism in psychology, such as Walter Mischel, and with situationism in philosophy, such as John Doris, do not deny the existence of all character traits.

Here would be a natural place to summarize the situationist movement, but fortunately there is no need, since it is already covered thoroughly in chapter 2 by our coeditors, William Fleeson and R. Michael Furr, and their colleagues.

Sixth Question: What Do Most of Our Character Traits Actually Look Like?

The interesting question, then, is *not* whether we have character traits but, rather, what these character traits tend to look like. Here, there is great deal of disagreement. Some of the leading answers include the following:

> Most people have global character traits understood as traditional virtues, such as honesty. These virtues are construed in Aristotelian terms as leading to stable behavior over time (i.e., telling the truth repeatedly in the courtroom) and consistent behavior across situations (i.e., at parties and in courtrooms). They involve a pattern of virtuous behavior, but also appropriately virtuous motivation as well.[35]
>
> Most people have global character traits understood as traditional vices, such as dishonesty. They involve stable and consistent behavior as well, but behavior that is only negatively valenced.
>
> Most people have local character traits, such as honesty in the courtroom. While these traits lead to stable behavior in the same nominal situations, they do not lead to cross-situationally consistent behavior across different nominal situations.[36]
>
> Most people have global character traits understood as Mixed Traits, such as a Mixed Helping Trait. These traits are stable and cross-situationally consistent, but the patterns of motivation and behavior to which they give rise are normatively positive in some respects and negative in others. They are not good enough to count as virtues but are not bad enough to count as vices.[37]

Trying to make progress on this debate is important from the perspective of the uses to which we put character traits (Fourth Question). For instance, we want to be able to predict how others will behave in morally important situations, and each of these hypotheses generates rather different predictions.

---

[35] For a version of this view, see Annas 2003.

[36] See Doris 1998: 507–508; 2002: 23–25, 64–66.

[37] See Miller 2013, 2014.

Seventh Question: How Can Good or Virtuous Character Traits Be Cultivated?

Naturally, we are not just interested in what character traits people have. Regardless of what we happen to actually be like, most of us are interested in developing a better character. But how can a good character be cultivated?

One source of controversy right off the bat is identifying who or what is primarily responsible for developing good character traits. In much of the philosophical and even the theological literature on virtue, responsibility for cultivating good character traits is placed squarely on the agent herself: individuals who regularly perform good actions develop good character traits, while those who regularly perform bad actions develop bad character traits. But many philosophers and theologians have also attributed a causal role to God. In the Christian tradition, for instance, central virtues are considered to be gifts of God rather than the product of human effort, and the Holy Spirit is thought to play an active role in the virtuous activity of religious believers, in addition to whatever contribution those believers themselves make. We will not say more about the role of God in character development here, as it is discussed at length in the chapters by Knobel (16), Yeo (24), and Cochran (25).

Leaving aside divine agency, there is also much philosophical and theological discussion of the relationship between religion and character development. Many philosophers and theologians have long argued that religion is essential to human flourishing, and that there are distinct virtues, such as piety, associated with worshiping God. More recently, however, scholars have also argued that religious practices can shape moral character more generally. In this volume, for instance, Terence Cuneo explores how Christian liturgical practices can shape one's narrative identity, while Elizabeth Bucar examines how Islamic veiling practices can shape one's view of modesty and humility.

There are also many different secular approaches to cultivating a good character. Here are a few examples which have been much discussed in recent years:[38]

(i) *Emulating models of good character*. The psychologists J. Philippe Rushton and Anne Campbell found that when a model signed up for a blood donation first, 18 of 27 participants did so as well, and 9 of them actually went through with the donation later, compared to none of the participants who actually donated blood in the no-model condition.[39]

[38] For more extensive discussion and references for these and other leading strategies, see Miller 2014: chapter 9.

[39] Rushton and Campbell 1977: 303. Similarly, Cashton Spivey and Steven Prentice-Dunn (1990) report that participants who, in person, saw a model donate money to an experimental assistant at a high level during 20 trials of a particular task, were themselves much more likely to do so at a comparatively high level.

(ii) *Increasing the salience of moral (and other) norms*. Participants in studies who were given an opportunity to cheat without any risk of getting caught typically did so to at least some extent. But in variants of these studies where the participants' moral norms were made salient to them, most did not cheat.[40]

(iii) *Educating people about the obstacles (sometimes unconscious or vastly underappreciated) to being virtuous*. In two studies Arthur Beaman and his colleagues had college students hear a social psychology lecture explaining the Latané and Darley model of group effects. The students were subsequently presented with a staged emergency—a victim of a bicycle accident in the one case, and a man sprawled against a wall in the other. Helping in the presence of a nonresponsive confederate was 67% versus 27% for controls in the first study, and 42.5% versus 25% in the second (even though in this study the helping opportunity occurred two weeks later than the lecture).[41]

(iv) *Seeking out situations which are conducive to morally admirable behavior, and actively avoiding situations which are conducive to immoral behavior*. In John Doris's well-known example, one should actively avoid a secluded dinner with a flirtatious colleague while one's spouse is out of town.[42]

Note that these strategies could be combined with each other, and they are by no means exhaustive of the possibilities which exist. Working out strategies for developing a good character is one of the most important areas in which further research on character is needed.

Clearly, there are many important and exciting questions pertaining to character that have not yet been fully answered. Collectively, the chapters in this volume help to advance our search for answers to all of them.[43]

References

Alston, William. (1975). "Traits, Consistency and Conceptual Alternatives for Personality Theory." *Journal for the Theory of Social Behaviour* 5: 17–48.

[40] For relevant discussion, see Mazar et al. 2008a, 2008b; and Shu et al. 2011.

[41] Beaman et al. 1978: 407–408, 410.

[42] Doris 2002: 147.

[43] This chapter draws on material from Miller 2013: chapter 1; 2014: chapter 1. Work on this chapter was made possible through the support of the Templeton World Charity Foundation. The opinions expressed are ours and do not necessarily reflect the views of the other editors of this volume, The Character Project, Wake Forest University, or the Templeton World Charity Foundation. For very helpful comments, we are grateful to Brandon Warmke.

Annas, Julia. (2003). "Virtue Ethics and Social Psychology." *A Priori* 2. Page references are to the version at http://apriorijournal.net/volume02/Annas1.pdf. Accessed August 15, 2011.

Badhwar, Neera. (2009). "The Milgram Experiments, Learned Helplessness, and Character Traits." *Journal of Ethics* 13: 257–289.

Beaman, A., P. Barnes, B. Klentz, and B. McQuirk. (1978). "Increasing Helping Rates through Information Dissemination: Teaching Pays." *Personality and Social Psychology Bulletin* 4: 406–411.

Bem, D., and A. Allen. (1974). "On Predicting Some of the People Some of the Time: The Search for Cross-Situational Consistencies in Behavior." *Psychological Review* 81: 506–520.

Brandt, Richard. (1970). "Traits of Character: A Conceptual Analysis." *American Philosophical Quarterly* 7: 23–37.

Buss, A. (1989). "Personality as Traits." *American Psychologist* 44: 1378–1388.

Buss, D., and K. Craik. (1983). "The Act Frequency Approach to Personality." *Psychological Review* 90: 105–126.

Butler, Douglas. (1988). "Character Traits in Explanation." *Philosophy and Phenomenological Research* 49: 215–238.

Doris, John. (1998). "Persons, Situations, and Virtue Ethics." *Noûs* 32: 504–530.

———. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Eaton, N., S. South, and R. Krueger. (2009). "The Cognitive-Affective Processing System (CAPS) Approach to Personality and the Concept of Personality Disorder: Integrating Clinical and Social-Cognitive Research." *Journal of Research in Personality* 43: 208–217.

Eklund, Matti. (2011). "What Are Thick Concepts?" *Canadian Journal of Philosophy* 41: 25–49.

Fleeson, W. (2001). "Toward a Structure- and Process-Integrated View of Personality: Traits as Density Distributions of States." *Journal of Personality and Social Psychology* 80: 1011–1027.

———. (2007). "Situation-Based Contingencies Underlying Trait-Content Manifestation in Behavior." *Journal of Personality* 75: 825–861.

Fleeson, W., and P. Gallagher. (2009). "The Implications of Big Five Standing for the Distribution of Trait Manifestation in Behavior: Fifteen Experience-Sampling Studies and a Meta-Analysis." *Journal of Personality and Social Psychology* 98: 1097–1114.

Fleeson, W., and E. Noftle. (2008a). "Where Does Personality Have Its Influence? A Supermatrix of Consistency Concepts." *Journal of Personality* 76: 1355–1385.

———. (2008b). "The End of the Person-Situation Debate: An Emerging Synthesis in the Answer to the Consistency Question." *Social and Personality Psychology Compass* 2: 1667–1684.

———. (2009). "In Favor of the Synthetic Resolution to the Person-Situation Debate." *Journal of Research in Personality* 43: 150–154.

Funder, D. (2008). "Persons, Situations, and Person-Situation Interactions." In *Handbook of Personality: Theory and Research*, 3rd ed., edited by O. John., R. Robins, and L. Pervin, 568–580. New York: Guilford.

Hampshire, Stuart. (1953). "Dispositions." *Analysis* 14: 5–11.

Heller, D., W. Perunovic, and D. Reichman. (2009). "The Future of Person-Situation Integration in the Interface between Traits and Goals: A Bottom-Up Framework." *Journal of Research in Personality* 43: 171–178.

Hogan, R. (1991). "Personality and Personality Measurement." In *Handbook of Industrial and Organizational Psychology*, 2nd ed., edited by M. D. Dunnette and L. M. Hough, 873–919. Palo Alto, CA: Consulting Psychologists Press.

———. (1996). "A Socioanalytic Perspective on the Five-Factor Model." In *The Five-Factor Model of Personality: Theoretical Perspectives*, edited by J. Wiggins, 163–179. New York: Guilford.

Hogan, R., C. DeSoto, and C. Solano. (1977). "Traits, Tests, and Personality Research." *American Psychologist* 32: 255–264.

Hursthouse, Rosalind. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.

Johnson, J. (1997). "Units of Analysis for the Description and Explanation of Personality." In *Handbook of Personality Psychology*, edited by Robert Hogan, John Johnson, and Stephen Briggs, 73–93. San Diego, CA: Academic Press.

———. (1999). "Persons in Situations: Distinguishing New Wine from Old Wine in New Bottles." *European Journal of Personality* 13: 443–453.

Kamtekar, Rachana. (2004). "Situationism and Virtue Ethics on the Content of Our Character." *Ethics* 114: 458–491.

Kupperman, Joel. (2009). "Virtue in Virtue Ethics." *Journal of Ethics* 13: 243–255.

Mazar, N., O. Amir, and D. Ariely. (2008a). "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research* 45: 633–644.

———. (2008b). "More Ways to Cheat: Expanding the Scope of Dishonesty." *Journal of Marketing Research* 45: 650–653.

Miller, Christian. (2013). *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.

———. (2014). *Character and Moral Psychology*. Oxford: Oxford University Press.

Mischel, W. (1968). *Personality and Assessment*. New York: John J. Wiley and Sons.

———. (1973). "Toward a Cognitive Social Learning Reconceptualization of Personality." *Psychological Review* 80: 252–283.

———. (1999). "Implications of Person-Situation Interaction: Getting Over the Field's Borderline Personality Disorder." *European Journal of Personality* 13: 455–461.

———. (2004). "Toward an Integrative Science of the Person." *Annual Review of Psychology* 55: 1–22.

Mischel, W., and Y. Shoda. (1995). "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure." *Psychological Review* 102: 246–268.

———. (1998). "Reconciling Processing Dynamics and Personality Dispositions." *Annual Review of Psychology* 49: 229–258.

Mumford, Stephen. (1998). *Dispositions*. Oxford: Oxford University Press.

Newman, L., and J. Uleman. (1989). "Spontaneous Trait Inference." In *Unintended Thought*, edited by J. Uleman and J. Bargh, 155–188. New York: Guilford.

Revelle, W. (1995). "Personality Processes." *Annual Review of Psychology* 46: 295–328.

Roberts, B. (2009). "Back to the Future: *Personality and Assessment* and Personality Development." *Journal of Research in Personality* 43: 137–145.

Rushton, J., and A. Campbell. (1977). "Modeling, Vicarious Reinforcement and Extraversion on Blood Donating in Adults: Immediate and Long-Term Effects." *European Journal of Social Psychology* 7: 297–306.

Russell, Daniel. (2009). *Practical Intelligence and the Virtues*. Oxford: Clarendon Press.

Shu, L., F. Gino, and M. Bazerman. (2011). "Dishonest Deed, Clear Conscience: When Cheating Leads to Moral Disengagement and Motivated Forgetting." *Personality and Social Psychology Bulletin* 37: 330–349.

Snow, Nancy. (2010). *Virtue as Social Intelligence: An Empirically Grounded Theory*. New York: Routledge.

Sosa, Ernest. (2009). "Situations against Virtues: The Situationist Attack on Virtue Theory." In *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, edited by C. Mantzavinos, 274–290. Cambridge: Cambridge University Press.

Spivey, C., and S. Prentice-Dunn. (1990). "Assessing the Directionality of Deindividuated Behavior: Effects of Deindividuation, Modeling, and Private Self-Consciousness on Aggressive and Prosocial Responses." *Basic and Applied Social Psychology* 11: 387–403.

Swanton, Christine. (2003). *Virtue Ethics: A Pluralistic View*. Oxford: Oxford University Press.

Upton, Candace. (2005). "A Contextual Account of Character Traits." *Philosophical Studies* 122: 133–151.

———. (2009). *Situational Traits of Character*. Lanham, MD: Rowman & Littlefield.

Wiggins, J. (1997). "In Defense of Traits." In *Handbook of Personality Psychology*, edited by. R. Hogan, J. Johnson, and S. Briggs, 95–115. San Diego, CA: Academic Press. Originally Presented in 1973.

Winter, D., O. John, A. Stewart, E. Klohnen, and L. Duncan. (1998). "Traits and Motives: Toward an Integration of Two Traditions in Personality Research." *Psychological Review* 105: 230–250.

Wright, J., and W. Mischel. (1987). "A Conditional Approach to Dispositional Constructs: The Local Predictability of Social Behavior." *Journal of Personality and Social Psychology* 53: 1159–1177.

2 }

Personality Science and the Foundations of Character

William Fleeson, R. Michael Furr,
Eranda Jayawickreme, Erik G. Helzer,
Anselma G. Hartley, and Peter Meindl

In the early twentieth century, *character* was considered central to the study of personality psychology. However, Hartshorne and May's (1928) influential study revealed that morally relevant behaviors, such as honesty, were less consistent than many researchers had expected. Observed consistency of single moral behaviors from situation to situation was modest, revealing correlations in the .2 to .3 range. Some psychologists interpreted these results to mean that there was no consistency in moral behavior, and thus that there were no stable, meaningful individual differences in moral behavior—in other words, character did not exist. Psychologists interested in individual differences subsequently labeled their efforts the study of "personality," rather than of character. The fallout was that researchers largely abandoned the study of ethics, morality, and virtue.

Our purpose in this chapter is to provide an empirical and theoretical foundation for the recent revival of character as an object of scientific psychological study. The study of personality has recently experienced an explosive regrowth. If character is related to personality more broadly, then the study of character may be able to take advantage of this movement, spurring a regrowth of its own. We begin by highlighting the character-relevant discoveries in personality more broadly. We then lay out five critical questions to guide researchers in the newly revived study of moral character and some possible answers we are pursuing.

We believe there are at least four reasons to investigate character. First, ethical behavior is a major goal in society and unethical behavior is a source of major societal problems, including crime, financial ruin, and pain. If

society can increase ethical behavior, it can also improve human relations, create a stable ground for investment and growth, and encourage human flourishing.

Second, much of moral behavior is likely based on long-standing psychological factors that exist within individual persons and on which individuals differ. However, the very claim that people possess global moral traits has been challenged; thus, it is important to determine whether individuals differ in moral character, how much they differ, how these differences manifest themselves in the individuals' lives, what the content is of individual differences in moral character, and how they function.

Third, character has grown in relevance to philosophical investigations of morality (see Miller and Knobel, chap. 1, this volume). In part, this is due to the revived interest in virtue ethics and the attendant focus on the psychological basis of character (Jayawickreme, Meindl, Helzer, Furr, & Fleeson, 2014). However, more broadly, empirical psychological studies of character have implications for many theories of ethics, and even for the proposition that behavior can be called ethical or unethical at all.

Fourth, character may be a primary lens through which people evaluate each other, and make decisions about whether to hire, associate with, or trust each other (Goodwin, Piazza, & Rozin, 2014).

Although *character* is a term intuitively imbued with value, its use has been surprisingly diverse. This diversity has been, not only across popular use, philosophical use, and psychological use, but also within psychological use. Within psychology, the field in which character is studied empirically, the different uses of character can be formed into classes, of which there are at least four (for a review, see Fleeson , Furr, Jayawickreme, Meindl, & Helzer, 2014).

We propose defining *character* as "characteristics that are descriptive of actions, cognitions, emotions, and motivations that are considered to be relevant to right and wrong according to a relevant moral standard" (Fleeson et al., 2014, p. 181).[1] This definition brings the complete package of character, morality, and virtue back to psychology, but leaves the ultimate question of what is moral open to conceptual debate. As such, this definition requires scholars to first stipulate a particular cluster of thoughts, feelings, or actions as being relevant to morality and then to study individual differences in that cluster (or disposition). For example, a researcher interested in studying compassion, honesty, or fairness (a fairly uncontroversial set of potential moral dispositions) would identify the thoughts, feelings, and actions that "count" as compassionate, honest, or fair; determine whether there are reliable individual differences in those characteristics; and eventually understand what makes some people more compassionate, for example, than others.

[1] We purposively take a more narrow definition of the term *character* by equating the word *character* with moral character.

We propose this definition of character for several reasons. First, it is in accord with the straightforward notion of positive character as the extent to which a person acts in morally admirable ways, but it goes beyond that notion. Second, this definition of "character" connects directly to a personality approach to the study of morality, which emphasizes characteristics of individuals. Finally, this definition allows leeway to researchers to designate the characteristics that they individually find moral. Research in this domain therefore has the ability to proceed in the face of eclectic definitions of morality: whether or not there exists disagreement about the definition of morality per se, there is at least agreement that these characteristics are important to study empirically.

The Fundamentals of Personality as a Basis for Character

In this chapter, we are calling for the study of moral character from the perspective of personality. That is, we believe researchers should study moral behavior as something that arises from long-lasting characteristics of individuals that usually and, at least potentially, differ between them.

In the last 20 years, the study of personality has experienced a renaissance. If character is related to normal personality, researchers can capitalize on this renaissance to make rapid progress in discovering the nature and functioning of character.

We want to acknowledge from the beginning that the chapter focuses on a trait approach to personality. There are other approaches to personality that are very important and that have great relevance to morality (McAdams & Olson, 2010; Roberts & Wood, 2006). For example, cognitive approaches focus on variables, such as life stories, scripts, beliefs, thinking styles, identity, and other thought-related characteristics (see Blasi, 1980, 1983; Gebauer, Riketta, Broemer, & Maio, 2008). Another approach, the motivational approach, focuses on the goals people pursue and their values. Although trait approaches include cognitive and motivational elements in their formulation, trait approaches do not give primary roles to cognition and motivation. Trait characteristics have received the most research attention in recent decades and are close to our definition of character. Thus, in this chapter we explore the relevance of trait characteristics to the study of character. We intend to explore other characteristics in future work (e.g., Helzer et al., under review; Fleeson et al., in prep.).

TRAIT CHARACTERISTICS OF INDIVIDUALS

Despite, or perhaps because of, people's intuitive grasp of the trait concept, there is surprisingly little consensus about the nature of traits. Most trait theories share the following five points, although each of these ideas may need to

be revised by further research and theorizing. First, traits describe qualities of individuals and emphasize the style in which individuals act, think, and feel. Consider the trait of boldness: a given individual may be high on the continuum of boldness, and therefore may act, think, and feel in a bold manner in response to many situations. Second, traits are characteristics on which people can differ from each other. For example, some individuals are bolder than others. Third, individuals most likely do not differ in whether or not they have a trait but rather in the degree to which they have a trait.[2] For example, although it may be convenient to speak of bold people and timid people, it is more likely that individuals differ along a continuous dimension of boldness. Fourth, traits endure for at least some extended time. For example, describing someone as a bold individual is useful only if that individual is bold for more than just a moment. Fifth, traits are broad descriptions of some kind of regularity, generality, or coherence in behavior, thought, and feeling. This generality may refer to the way the individuals act across different situations, to the way they act in significant, defining situations, or to a wide range of ways they act. For example, bolder individuals engage in a variety of bold actions, such as stating opinions, taking risks, and making decisions.

THE PERSON-SITUATION DEBATE

A major endeavor in personality psychology, persisting for most of the twentieth and into the twenty-first century, has been investigating whether personality exists in sufficient breadth and strength to be worth studying. In recent years, this debate has spread to philosophy, where the existence of character (understood as we previously defined it) has been plausibly challenged. Interestingly, although this debate has propelled much of the research supporting the existence and importance of "normal" personality traits, original doubts about personality stemmed from doubts about moral character.

In the early part of the twentieth century, researchers were content to study behavior, and in particular moral behavior, as the result of dispositions. This position, which sees the person as the source of many important behavior regularities, was the compelling lay view that traits do exist, because behaviors are consistent. Because it appeared self-evident that people do behave similarly to themselves and differently from others, it was assumed that there must be underlying traits that cause those differences.

This position was challenged throughout much of the middle part of the twentieth century with the rise of situationism. Situationists pointed to three classes of evidence to cast doubt on the usefulness of a person approach to the study of behavior—an approach that assumes that behavior variability

[2] We recognize this view is not shared by some philosophers who adopt a binary perspective in which individuals either possess a trait or are absent of that trait (see Miller, 2013).

arises from variability in people's characteristics. The first type of evidence involved cross-situational correlations. Cross-situational correlations reveal the degree to which individual differences in behavior in one situation are repeated in a new, different situation. For example, a high cross-situational correlation would mean that those individuals who were most (and least) honest when taking a test would also be most (and least) honest when talking with their peers. In a classic study of dispositional honesty, Hartshorne and May (1928) found that the correlation between children's honesty behavior in any two situations was far from 1.00 (perfect consistency)—typically, around .20. Subsequent reviews of the literature repeated this lesson, reporting that cross-situational correlations almost never rose above .40 (e.g., Hunt, 1965; Mischel, 1968; Vernon, 1964; Wallace, 1966). Thus, a person's honest or dishonest behavior in one situation was viewed as only a weak indicator of their honesty in another situation—so-called honest people frequently could (and did) act dishonestly, and so-called dishonest people could (and did) act honestly. This evidence suggested that individuals may not have a strongly consistent way of acting and that therefore traits are not powerful, if they exist at all.

The second type of evidence was the documentation of powerful and pervasive effects of situations on people's behavior (for a review, see Miller, 2013). Many situationists, including Ross and Nisbett (1991), pointed to the counterintuitive and robust ways in which situations had effects on behavior. For example, Latané and Darley (1970) found that average participants sought help for someone apparently having an epileptic seizure after only 52 seconds, but that the mere presence of a few bystanders increased the help-seeking delay to nearly three minutes. Such large effects of apparently small changes in the situation suggested that situations are much more powerful in affecting moral behavior than any traits could be.

The third type of evidence included results showing that mental biases lead people to overestimate the actual degree of consistency in a person's behavior. Several studies demonstrated that individuals were insensitive to situational influences on people's behavior and stubbornly persisted in their dispositional inferences, even when strong situational forces made such inferences unlikely (Jones & Harris, 1967; Nisbett & Ross, 1980; Ross, 1977). This raised the damning possibility that "traits" might only be a concept that existed inside the heads of perceivers. Actual dispositions, such as honesty or compassion, were simply mental illusions.

These situationist arguments were persuasive to researchers at the time. Within psychology, many departments stopped hiring personality psychologists (Swann & Seyle, 2005), and did away with studying the role of the person and his or her history, beliefs, and values in the creation of moral behavior, flourishing, and growth. Within moral philosophy, a group of philosophers (Harman, 1999; Doris, 1998, 2002) took the research discussed earlier as

evidence against the existence of virtues, thus claiming that the investigation of virtue ethics was empirically unjustified. As Doris (2002) puts it:

> The experimental record suggests that situational factors are often better predictors of behavior than personal factors . . . In very many situations, it looks as though personality is less than robustly determinative of behaviors. To put things crudely, people typically lack character (p. 2).

THE RENAISSANCE OF PERSONALITY

In response to the situationist argument, personality psychologists refocused their work on accumulating a mass of empirical evidence for the existence and usefulness of trait concepts. Thus far, this work has been mostly limited to the study of non-moral traits, but it has been very fruitful. There is now considerable consensus about the general structure and hierarchical format of individual differences in traits across several cultures (Goldberg, 1993; but see Ashton & Lee, 2007; Benet & Waller, 1995; Clark & Watson, 1999, for other structures). Traits have been found to be relatively stable across adulthood (Block, 1971; Roberts & DelVecchio, 2000). Behavioral genetics research has shown that broad personality traits have a genetic component as well (Krueger, South, Johnson, & Iacono, 2008), and that heritability is never zero (Turkheimer, 2000). Further, contrary to what situationists proposed, personality is not just in the "eye of the beholder": individuals can accurately judge their own and others' personalities (Kenrick & Funder, 1988). Several studies have demonstrated that targets and observers also agree in their judgments about the targets' personalities, including both close others and new acquaintances (Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008). Collectively, these findings remove doubts about the existence of powerful and enduring personality traits, and raise important questions about the nature of these underlying dispositions.

Perhaps the most powerful finding to come out of personality research in the past decade is that traits can predict important life outcomes, both concurrently and prospectively over decades (Kern, Duckworth, Urzua, Loeber, Stouthamer-Loeber, & Lynam, 2013; Ozer & Benet-Martinez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Extraversion and neuroticism, for example, predict positive affect and negative affect, respectively, at about the .3 to .4 level (Lucas & Fujita, 2000; Costa & McCrae, 1980), which is greater than any other variable's ability to predict affect. This effect has now been replicated across multiple labs, age groups, and cultures (Lucas & Fujita, 2000). It has even been demonstrated within introverts, as Fleeson, Malanos, & Achille's (2002) findings revealed that both extraverts and introverts were happier when they acted extraverted, and that manipulating state extraversion induced state positive affect (McNiel, Lowman, & Fleeson, 2010).

Other traits have been shown to predict more objective life outcomes. Several studies have found that children higher in conscientiousness have better health, job performance, and romantic relationships and show less antisocial behavior in adulthood (Hampson, Edmonds, Goldberg, Dubanoski, & Hillier, 2013; Shiner, Masten, & Roberts, 2003). Conscientiousness predicts work and school performance about equally as strongly as does intelligence (Noftle & Robins, 2007; Ones, Viswesvaran, & Schmidt, 1993), regardless of type of measure of work performance. Physical health and longevity are influenced by personality: People higher in conscientiousness, extraversion, emotional stability, and agreeableness live longer. These links are strong, and equal to or greater than those connecting socioeconomic status with mortality. Personality traits are also directly related to numerous highly prevalent mental disorders with substantial documented social costs, including mood, anxiety, substance use, and antisocial disorders (e.g., see Kern et al., 2013; Stautz & Cooper, 2013).

This explosion of personality research has made great strides in advancing our knowledge of traits, individual differences, predicted life outcomes, and personality change. We believe this research sets the stage for the empirical study of moral character. Just as researchers have done for so-called normal personality, researchers of character must establish the existence of robust, stable individual differences in moral behavior that persist across time and situations, and that carry consequences for important life outcomes. This first step is essential, given that situationism, the perspective whereby moral behavior results from impersonal and often fleeting situational forces, is so prominent in moral psychology.

Five Research Questions on Character

We propose five research questions that we believe will ignite the explosion in the study of moral character.

QUESTION 1: WHAT IS THE RELATIONSHIP OF MORAL BEHAVIOR TO PERSONALITY TRAITS IN GENERAL?

To use personality research to inform research on moral character, there should be a strong connection of morality to "normal" personality traits. The connection of morality to personality traits in general, however, is not yet known.

There are at least five possible connections. One possibility is that existing personality traits predict individuals' levels of moral character. On this possibility, character is a separate domain of personality, but one that is tied to other personality characteristics in a systematic way. Recent research in this area suggests that there is at least a strong association between personality traits such as agreeableness and conscientiousness and self-reported moral

attributes (see Noftle, chap. 22, this volume), but it remains unknown how personality might predict specific aspects of moral character.

A second possibility is that moral character is not a separate domain but is rather part of the known trait hierarchy, the Big Five (agreeableness, extraversion, conscientiousness, openness, and neuroticism). The Big Five purportedly cover all personality characteristics that are important (John et al., 2008), and moral character is presumably an important personality characteristic, implying that moral character is already covered in the Big Five. Moral character could be part of a single trait within the Big Five, or it could be divided and contained in multiple traits of the Big Five. For example, following rules could be part of conscientiousness, whereas standing up to authority could be part of extraversion.

A third possibility is that moral character is contained within the recently discovered sixth factor of the HEXACO model—namely, humility/honesty (Ashton & Lee, 2007). Honesty is a prototypical aspect of morality, but the trait may not cover other aspects of morality.

A fourth possibility is that moral character is a trait, or set of traits, but is not part of the Big Five and HEXACO. There are some who have argued that moral character may consist of several major domains, similar in structure to the Big Five (see Miller, 2013).

The fifth possibility is that moral character is not a trait but rather is a cognitive or motivational characteristic (Lapsley & Narvaez, 2004). That is, morality could describe individual differences in people's cognitive styles and identities, or their motives and goals.

Psychologists cannot yet say which of these possibilities is true, but it is critical to discover which one is true in order to identify what kind of characteristic moral character is, and also to apply existing knowledge about personality to moral character.

QUESTION 2: IS MORAL BEHAVIOR CONSISTENT? DO BROAD MORAL TRAITS EXIST?

The impressive evidence in favor of broad traits notwithstanding, the question about the viability of moral dispositions in particular remains open. Although nearly 90 years have passed since Hartshorne and May (1928) published their findings, and though there is now strong evidence in support of broad traits such as extraversion, neuroticism, agreeableness, conscientiousness, and openness, only recently has evidence emerged regarding the claim that moral behavior is a feature of moral or immoral situations, and not people.

It is essential to the study of character that moral behavior be consistent in some form or other. Otherwise, there will be no meaningful individual differences regarding moral behavior to study. Thus, the question of whether moral behavior is consistent is perhaps the most pressing for research in this domain.

Demonstrating consistency in moral behavior would establish that individuals vary meaningfully in their morality styles: their patterns of moral behavior in response to eliciting social situations. The challenge for the study of moral character, therefore, is to reconcile the proposed existence of broad moral character traits with established demonstrations of the variability in moral behavior. How can robust individual differences in honesty or compassion exist if honest people sometimes act dishonestly and compassionate people sometimes act cruelly?

Understanding Consistency in Non-Moral Traits

For non-moral traits, the density distributions model was offered as a way to reconcile how robust individual differences in traits can exist with the empirical fact that people are highly variable in how they act (Fleeson, 2001). It is possible that the density-distributions approach can also apply to moral traits. This approach starts with the concept of a "personality state" (Cattell, Cattell, & Rhymer, 1947; Fleeson, 2001; Fleeson, Malanos, & Achille, 2002; Fridhandler, 1986). A state is defined as having the same content as a corresponding trait but applying for a shorter duration. For example, an extraverted state has the same content as extraversion (talkativeness, energy, boldness, assertiveness, etc.), but applies as an accurate description for only a few minutes to a few hours rather than for the months or years that a trait description applies. Although the word "behavior" is used to refer to the state, this use is meant in the broad sense, emphasizing how the individual is acting but also including how the individual is feeling and thinking at the moment. The key is that the state concept assesses what the individual is actually doing, at the moment he or she is doing it, in real situations.

In the density-distribution approach, personality state manifestation is measured across many different points in a person's day or week and the frequency with which the individual acted at each level of the state dimension is recorded. The accumulated frequencies are then combined to form a distribution for each individual of each trait. The distribution provides a comprehensive account of the individual's behaviors across several days. In the moral domain, a moral "state" would be the degree to which the person is being moral at a given moment. A moral density distribution would be the accumulation of the individual's moral behaviors across a period of time.

Consistent with the findings of social psychology, density distributions revealed non-moral behavior to be far less consistent than expected. *Within-person variability* is defined as the degree to which a single individual acts differently on different occasions. It is possible to calculate a standard deviation for each person, separately, describing the amount that his or her behavior differed across the occasions of a period of observation. A large standard deviation means that the individual's state levels varied widely, so that on some occasions he or she behaved at a high level of the state, and on other

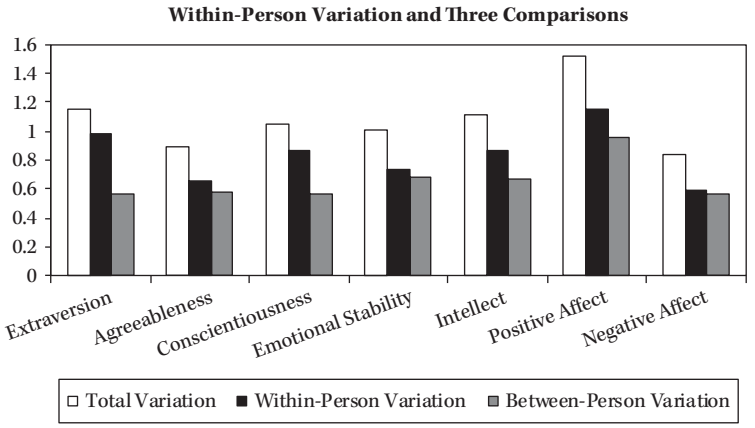

FIGURE 2.1 Within-Person Variation and Three Comparisons.

Source: Adapted from Fleeson, 2001, with permission.

occasions he or she behaved at a low level of the state. A low standard deviation, in contrast, indicates that the individual behaved the same or similar way on most occasions. The average of the standard deviations across all people is the typical individual's amount of variability. Figure 2.1 shows that the amount that one typical individual varied in his or her behavior across two weeks was (1) about the same as the amount of within-person variation in affect, something that is commonly known to vary so much that it is hard to conceive of it as a trait; and (2) more than the amount of variability between individuals, meaning that individuals differ from themselves more than they differ from others.

We propose that moral behavior may be equally variable within persons. That is, each individual will differ in the morality of his or her actions quite a bit. Some types of moral behavior may have a wider range, such as lying, and other types of moral behavior may have a narrower range, such as stealing. Thus, it is an important empirical question to determine which moral behaviors show the same surprising degree of sizeable within-person variability that personality behaviors in general do. This is the kind of variability in behavior that is behind the situationist challenge to the existence of character—the data we have collected is consistent with the data on which the situationists base their position.

Relatedly, individuals with different traits overlap considerably in their behavior. Fleeson and Gallagher (2009) conducted a meta-analysis of 15 studies conducted over eight years to determine the implications of traits for distributions of behavior. Participants were categorized as extraverts or introverts on the basis of a standard retrospective summary questionnaire. In figure 2.2, the gray distribution is the typical behavior distribution of a very extraverted individual; the unshaded distribution is the distribution of a very introverted

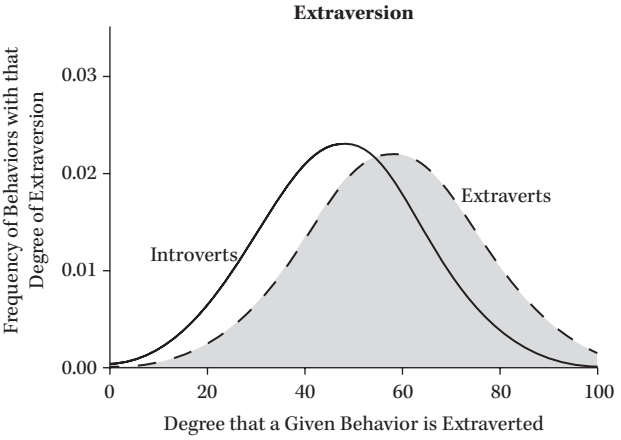

FIGURE 2.2 Distributions of actual trait-manifestations in behavior for those high and those low on a trait.

Source: Adapted with permission from Fleeson & Gallagher, 2009.

individual. It can be seen that extraverts and introverts do enact different behaviors, because the distributions do not overlap completely. However, there is also a great deal of overlap in how they act: Extraverts quite regularly act introverted and introverts quite regularly act extraverted. The difference between extraverts and introverts is not in the way they act, or in the extremes, but in the marginal increased frequencies with which they enact mid-range extraverted and introverted behaviors.

Moral behavior may show a similar overlap. Just as the amount of overlap in behavior between individuals with different trait levels was surprisingly high, the amount of overlap in moral behavior between individuals with different character levels may be surprisingly high. Thus, it may be the case that individuals who are truly different from one another in traits such as honesty or compassion, for example, act similarly to one another much of the time. It is imperative to obtain direct, empirical accounts of moral behavior rather than relying on retrospective summary questionnaires that may overstate the distinctions between those with different levels of character.

Nonetheless, individual differences in average behavior are highly stable. Sizeable within-person variability and high overlap appear to imply that people don't reliably differ from each other in their behavior (Mischel, 1968; Ross & Nisbett, 1991). To investigate whether there were nonetheless stable individual differences for non-moral behaviors, we examined the distributions as wholes rather than as specific, single behaviors.

Although each individual varied considerably, each individual also had a central point or tendency around which he or she varied (Diener & Larsen, 1984; Epstein, 1979; Fleeson & Noftle, 2008b; Moskowitz, 1982). The question was whether each individual's central point remained stationary and

therefore indicated an essential part of that individual's behavior. By splitting the data in half, we were able to calculate two central tendencies for each individual and then correlate them. The extremely high correlations (between .71 and .87) meant that differences between individuals in their average tendencies were highly stable and highly predictable (Fleeson, 2001). This is the key evidence against the situationist position—differences between individuals in their average behavior were as consistent and predictive as any variable in psychology. The meta-analysis of 15 studies found that retrospective questionnaire measures of traits predicted these aggregated means at about the .5 level (Fleeson & Gallagher, 2009). Thus, personality differences show up in averages, rather than in qualitative differences in the single behaviors they enact. Similarly, character may become most clear only in averages of moral behavior rather than in qualitative differences in single moral actions.

Does Moral Behavior Follow Density Distributions?

We propose that moral behavior may also follow density distributions. Moral behavior may be highly variable within persons, so that the same person may act morally on some occasions and less morally on others. People may overlap considerably in the morality of their behaviors over time. Nonetheless, individuals may differ very reliably in the distributions as wholes, such that some people have distributions at a higher level of moral behavior than others.

Consider honesty. The density distributions approach would start with the notion that people are honest sometimes and dishonest other times. Over time, each person accumulates a distribution of honesty. The distribution would depict the number of times the person was highly honest, moderately honest, moderately dishonest, and highly dishonest, including finer gradations. These distributions might be wide, with lots of within-person variability. Different people's distributions might overlap with each other. Nonetheless, some people would have distributions that shifted more to the honest side than would others, and these differences in distributions would stay highly stable. Thus, there is a great deal of consistency in the overall distribution of honest behavior, and individual differences in honesty exist.

QUESTION 3: WHAT IS THE INTERNAL CAUSAL STRUCTURE PRODUCING CHARACTER?

If individuals do differ in the morality of their behavior, it becomes paramount to figure out why they differ. This means explaining both why people differ from each other and why a given person acts morally sometimes and less morally at other times. Thus, we need to identify the causal determinants of intra- and inter-individual differences in moral behavior.

Large within-person variability is interesting on its own for elucidating the nature of moral character or personality in general. But it can also be harnessed as a way to explain moral character or personality in general.

Identifying the Contingencies of Moral Behavior

Research on normal personality has begun to explain traits. For example, in one study, participants completed standard experience-sampling method (ESM) reports: a form of diary-method data collection in which participants provide ratings in real time of their experiences at multiple points during their day. Participants provided ESM reports across 35 different time points, but also rated the degree to which several situation features were present at the time of the report. Hierarchical linear-modeling analyses produced betas indicating how much each subject's expression of each trait covaried with the occurrence of each situation feature.

The results showed that situation features significantly predicted trait expression. As shown in figure 2.3, these results indicate successful identification of friendliness of others (dark bars), task-orientation (white bars), and anonymity (gray bars) as meaningful situation features for extraversion, agreeableness, and conscientiousness. Thus, friendliness of others may be a potent factor in producing extraversion; becoming extraverted when others become friendly may be an underlying mechanism of extraversion. Fournier, Moskowitz, and Zuroff (2009) conducted a similar analysis, but used the traits of the interpersonal circumplex and if-then situation behavior profiles, and found

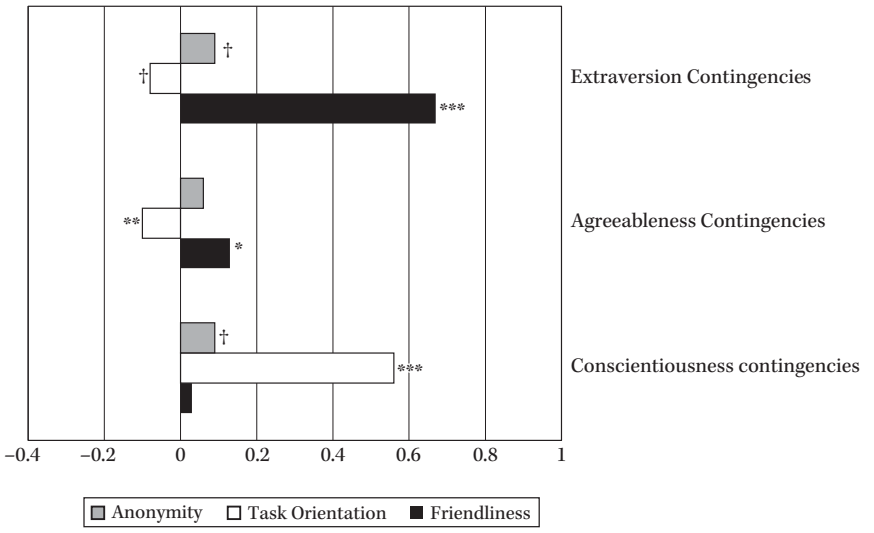

FIGURE 2.3 Magnitude of situational contingencies for extraversion, agreeableness, and conscientiousness.

Source: Fleeson, 2007.

individual differences in the organization of behavior that were not revealed by traits alone. We term these contingent relationships between eliciting social situations and people's trait-relevant behavior (e.g., the friendliness of others elicits extraversion) *contingencies*—and propose that they will be important mechanisms underlying moral character as well. Contingencies likely explain the occurrence of moral behavior, the functioning of moral behavior traits, and individual differences in moral behavior.

Individualized Contingencies

A particularly important finding from this research was that there were significant individual differences in the magnitude of contingencies for five of the nine possible cases. That is, different people responded in different ways to the same situations, and these individual differences were reliable (Fleeson, 2007). We propose that similar individual differences in moral behavior contingencies will exist as well. The empirical record, in fact, already suggests that this is true. In classic situationist studies of moral behavior, there are always some people in each condition who will act in a way that is contrary to prediction: Although a sizeable number of people in Darley and Batson's (1973) study didn't help when they were in a hurry, some did. This suggests that different people respond differently to situational triggers of moral behavior. Understanding these individual differences in situational contingencies will be essential to the study of character.

What Other Mechanisms Explain Moral Behavior?

In addition to contingencies, there are other likely explanations for trait-relevant behavior. For example, *social-cognitive variables* are psychological constructs, such as beliefs, expectancies, goals, plans, and emotions (Mischel, 1973). Whole trait theory (Fleeson & Jayawickreme, in press) proposes that the explanations of traits consist primarily of links between a broad set of social-cognitive variables. Social-cognitive variables are linked together in the person's mind, such that some of these variables lead to others. The social-cognitive variables are then linked to state outcomes. Similarly, Narvaez, Lapsley, Hagele, and Lasky (2006); Snow (2009); Aquino, Freeman, Reed, Lim, and Felps (2009); and Miller (2013) have adopted a social-cognitive perspective of moral character as a complex knowledge structure, in which individuals' moral components are activated in salient social situations that they encounter. For some individuals (e.g., those high in virtue), these moral constructs will be readily available when they respond to social information, causing them to act morally; whereas other individuals will have other non-moral constructs available.

Other research supports social-cognitive explanations of traits as well. For example, Bleidorn (2009) and Heller, Komar, and Lee (2007) demonstrated that the mechanisms underlying the Big Five were motivation-based. Conway, Rogelberg, and Pitts (2009) used ESM to provide evidence for the affect-based

mechanisms underlying agreeableness behavior. Gebauer et al. (2008) has shown that belief in genuine altruism is an important predictor of moral behavior.

In addition to these social-cognitive variables, researchers are beginning to uncover biological explanations of personality traits and moral behavior. Twin studies examining personality development over time suggest that genetic factors may be responsible for the stability of personality traits in young adulthood, while changes in personality can be explained by environmental factors (Kandler, Bleidorn, Riemann, Spinath, Thiel, & Angleitner, 2010; see also Hopwood, Donnellan, & Blonigen, 2011). Although still nascent, recent research has linked personality traits and behaviors with specific patterns of neurological activity and levels of neurobiological substrates (see DeYoung, 2010). For example, two recent studies demonstrate that people high in narcissism show activity in the social pain network region of the brain when feeling socially excluded (Cascio, Konrath, & Falk, 2014). This line of research also indicates that biological correlates may play an important role in specific dimensions of moral character as well, such as empathy and altruism (Ho, Konrath, Brown, & Swain, 2014).

Clearly, both social-cognitive variables and biological variables may be part of the mechanisms producing individual differences in moral behaviors. It will be important for future research in this area to identify the different underlying causal factors involved in moral behavior, whether they be social-cognitive, biological, or situation-behavior contingencies.

QUESTION 4: THE ROLE OF MORAL CHARACTER IN SELF-KNOWLEDGE AND OTHER KNOWLEDGE

The next set of questions concerns whether moral character plays an important role in individuals' understanding of themselves and other people. First, to what degree does an individual's beliefs about his or her own moral character match other people's beliefs about his or her character? For example, if a person perceives herself as altruistic, do other people also perceive her as altruistic? And to what degree do acquaintances from different domains of a person's life agree about the person's character? Second, are the beliefs a person possesses about an individual's moral character particularly potent in shaping his or her understanding and evaluation of that individual, trumping perceptions about other personality traits, such as the degree to which a person is intelligent or sociable? Finally, do individuals differ in the extent to which their interpersonal understanding and evaluations are shaped by their beliefs about others' moral character?

Answering these questions will provide insights into the robustness of moral character traits and the role of moral character in both intrapersonal and interpersonal dynamics. If different people agree in their beliefs about

or perceptions of individuals' moral character, it suggests that character is indeed a stable quality of individuals. That is, if two observers come to the same conclusion, without talking to each other, about an individual's level of morality, it suggests (though certainly does not prove) that there is something meaningful and real about that individual's moral behavior that persists across time and different situations. Such interpersonal agreement also signals that the perception of moral character is not an error or an illusion, but the product of a reality shared between different observers. If moral character traits are particularly potent in interpersonal perception and evaluation, it likely means that people are highly motivated to know others' character, and are perhaps perceptually "tuned in" to information that reveals character. Moreover, such findings would suggest that moral character traits have particularly robust effects on one's self-concept and on one's interpersonal relationships.

These questions have been examined extensively by studying the context of "normal" personality traits, providing important insights into the existence, nature, and interpersonal importance of those traits. The insights gleaned from this research can be used as the basis for setting an agenda for the study of moral character.

Correspondence between Self-Perception and Others' Perceptions

For more than a century, psychologists and sociologists examined the correspondence between perceptions of one's own personality and other peoples' perceptions of one's personality (e.g., Cooley, 1902; Mead, 1934; Shrauger & Schoeneman, 1979; Funder, 1980; Vazire, 2010). Although social psychological research in the 1950s, 1960s, and 1970s emphasized errors that people make when perceiving themselves and others, more recent research reveals that people are, in fact, more accurate in their perceptions of themselves and others than was previously believed (Funder, 1987, 1995).

A representative example of research revealing correspondence between self-perceptions and others' perceptions is provided by Furr, Dougherty, Marsh, and Mathias. (2007). For this research, 98 adolescents rated themselves on 35 fundamental personality traits (e.g., assertiveness, anxiety, competence), and their mothers rated the adolescents on the same traits. For each trait, the correspondence between self-ratings and mother ratings was reflected by the "self-other agreement" correlations between adolescents' self-ratings and their mothers' ratings. High positive correlations (between .20 and +1) reflected good correspondence; whereas a zero or negative correlation (between 0 and −1) reflected poor correspondence. As shown in figure 2.4, correlations were greater than zero for all 35 traits, and, in fact, significantly greater than zero (and greater than .20) for 34 traits. The average correlation was .42, indicating substantial agreement between mothers and their adolescent children about the children's personality traits.

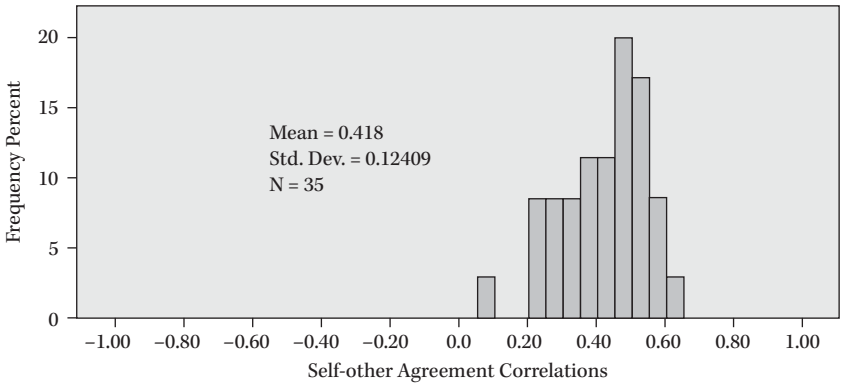

FIGURE 2.4 Distribution of self-other agreement correlations.

Source: Adapted with permission From Furr, Dougherty, Marsh, & Mathias, 2007.

Substantial Correspondence among Others' Perceptions of an Individual, for at Least Some Traits

An issue closely related to self-other agreement or correspondence is the correspondence between two people's perception of a third person. Often called "interjudge consensus," this reflects the degree to which people agree on an individual's reputation. Previous research reveals high levels of consensus for some core traits.

Consider the following findings (see figure 2.5), based on data reported by Carlson and Furr (2009). In this study, each of 101 "target" participants was described by (up to) six acquaintances who knew him or her well—two parents, two hometown friends, and two college friends. Acquaintances rated each target in terms of the Big Five personality traits, and analyses reveal robust agreement between acquaintances' ratings of their targets for each trait. For example, the correlation between mothers' ratings and fathers' ratings of their target's agreeableness was a sizeable .57, and the correlation between the two hometown friends' ratings of the targets' conscientiousness was an even larger .65 (see figure 2.5A for all five traits). Such large positive correlations indicate that people develop common perceptions of individuals.

Although consensus is strong among acquaintances from the same social context, what happens when acquaintances are from different social contexts? Because they see their child in the same (or quite similar) social contexts, a mother and father might have highly similar views of their child. But to what degree do a target's mother and his or her college roommate view the target in similar ways? The mother and the roommate likely see the target in dramatically different social contexts, potentially seeing dramatically different behaviors and thus potentially forming substantially different perceptions of the target's personality.

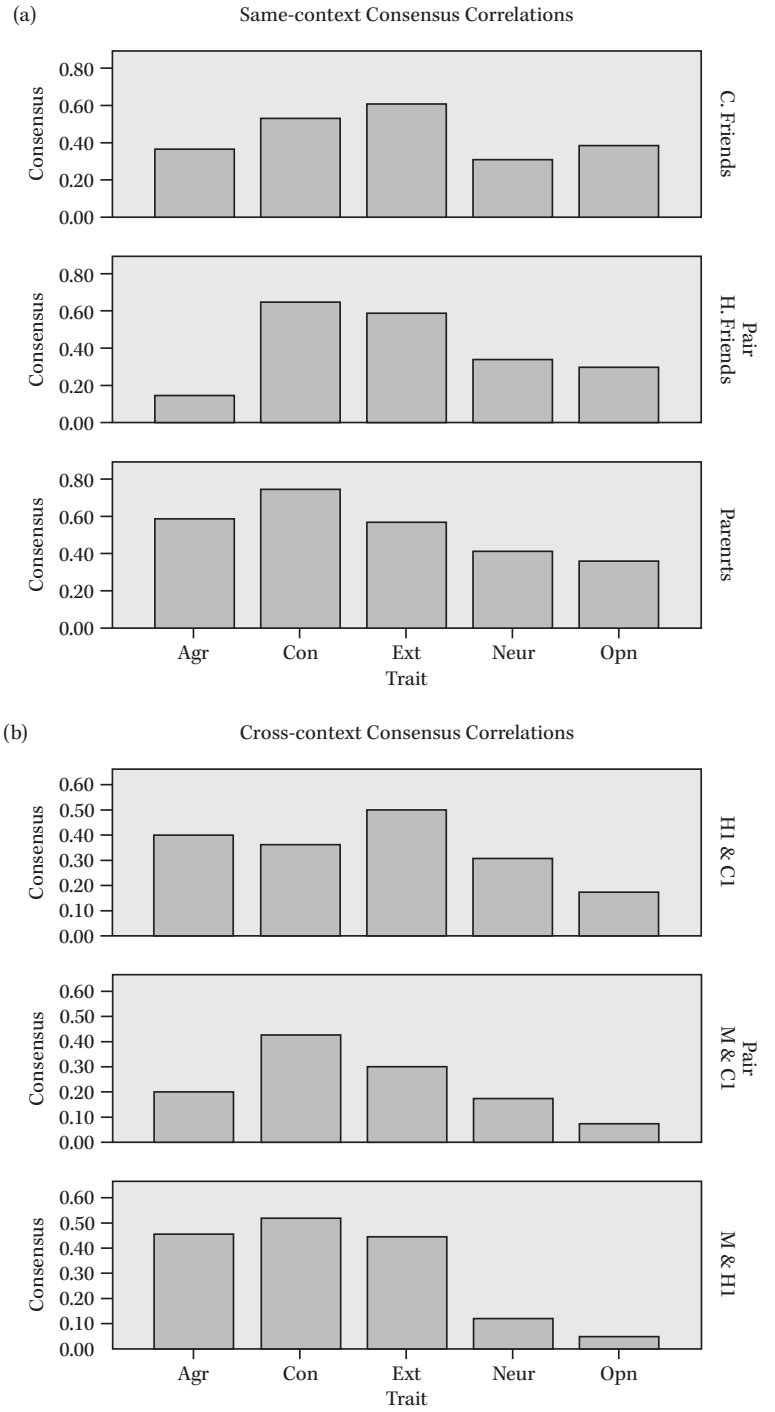

FIGURE 2.5 A) Same-Context Consensus Correlations. B) Cross-Context Consensus Correlations.

Source: Based on data from Carslon & Furr, 2009.

Consensus between people from different social contexts is indeed generally lower than consensus between people from the same social context; however, it generally remains quite robust. For example, the correlation between mothers' ratings and a hometown friend's ratings of targets' agreeableness was .46, and the correlation between a hometown friend's ratings and a college friend's ratings of conscientiousness was .37 in Carlson and Furr's (2009) data. As shown in figure 2.5B, cross-context correlations are generally positive and statistically significant, particularly for agreeableness, conscientiousness, and extraversion.

Considering the general pattern of findings regarding self-other agreement and interjudge consensus, evidence suggests that one's behavioral style is in fact stable enough to produce meaningfully similar impressions in people from a variety of social contexts. We believe that similar results will emerge for moral character traits, and, indeed, recent findings support this. Two recent papers reveal significant self-other agreement and interjudge consensus in the perception of several moral character traits, general moral evaluations, and morally relevant emotions of guilt and shame (Cohen, Panter, Turan, Morse, & Kim, 2013; Helzer, Furr, Hawkins, Barranti, Blackie, & Fleeson, in press). However, such studies are just beginning, and we hope that research expands on these findings in important directions.

Some traits can operate as lenses through which general perceptions are filtered.

Previous research shows that, in the process of forming impressions of others, some traits seem to be filters or lenses that color one's overall perception and evaluation. Moral traits, because of their putative importance, might be lenses for many or most individuals; but it is also possible that people differ in the extent to which their perceptions are colored by information about moral traits. Intuition suggests that some people are more moralistic than others, and it is not yet clear whether there is a general tendency, or individual differences in the tendency, to preferentially perceive and interpret information about moral traits in judgments of self and others.

Perhaps the most well-known demonstrations of traits as lenses or filters are Asch's (1946) classic studies of impression formation. Asch's participants read a series of trait terms reflecting a hypothetical "target" person's personality, and they were asked to "form an impression of the kind of person" described by those terms. For example, some participants read the terms "energetic," "assured," "talkative," "cold," "ironical, "inquisitive," and "persuasive," while other participants read the same terms, with the sole exception that "cold" was replaced by "warm." Reflecting their general impressions of the target person, participants rated that person on additional personality characteristics (e.g., generous vs. ungenerous, shrewd vs. wise, humorous vs. humorless, ruthless vs. humane, restrained vs. talkative). Asch examined differences in overall

impressions that were formed by the participants who believed the target to be warm and those who believed the target to be cold.

As shown in figure 2.6, results revealed that the warm versus cold difference affected a variety of perceptions—for example, participants viewed the "warm" target as more generous, wise, happy, humorous, humane, and imaginative than the "cold" target. Importantly, Asch (1946) found that alternative distinctions—e.g., between polite and blunt—produced weaker effects on overall impressions than did the warm-cold distinction. These results reveal that some trait dimensions (e.g., warm-cold) are more central to overall impressions than are other dimensions (e.g., polite-blunt).

Overall, Asch's (1946) experiments produced important insights into the processes through which people form impressions. These include (a) the possibility that some traits are more potent in shaping impressions than are other traits and (b) the possibility that a particularly potent trait can affect the way other traits are interpreted (e.g., a person who is warm and intelligent is viewed as "wise" while a person who is cold and intelligent is viewed as "shrewd" or "calculating").

At least two fundamental issues were beyond the scope of Asch's work, but are crucial for a deeper understanding of moral character's impact on self-knowledge and other-knowledge. First, his work did not explore the effects of moral character specifically. Interestingly, this lack of specific attention to moral character in empirical studies of impression formation persisted for several decades. Recent research, however, clearly reveals the strong and likely dominant effect that moral character has on peoples' evaluations of themselves and others (Goodwin, Piazza, & Rozin, 2014; Strohminger & Nichols, 2014). Second, Asch did not examine the possibility that people differed in

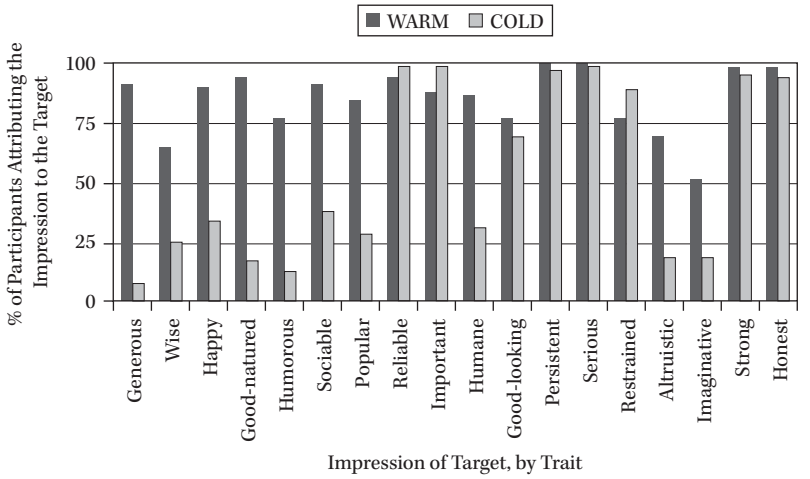

FIGURE 2.6 Percentage of Participants Attributing Impression to Target, by Condition.

the potency with which information about moral traits affects understanding and evaluation of others. Some people's perception of themselves and others might be heavily affected by information about moral character traits, while other people's perceptions are less affected by such "moral information." Such differences could have deep and far-reaching effects on intrapersonal and interpersonal dynamics.

QUESTION 5: THE DEVELOPMENT OF MORAL CHARACTER

Our final set of questions regarding the study of character focus on the development of moral character throughout the lifespan. How early does moral character develop in life, and continue to develop throughout childhood, adolescence, and adulthood? Does moral character remain stable over time, or show variability? How does the character of an exemplary moral individual differ from other individuals? How can we study character development from a personality science perspective that recognizes the variability of people's moral behaviors across social situations and contexts, yet captures the stability of their moral character over time? Research focused on answering these questions holds the potential to provide insight on how to promote the development of moral behavior, improve social functioning, and prevent antisocial behavior.

Research on moral development has benefited from increased progress in recent decades than has research on moral personality. Thus, this topic starts at a more developed level than does the research on character in general. Nonetheless, many opportunities for advances are still available.

Moral Character Is Established Early in Development

Past work suggests that moral character is established early and continues to develop throughout childhood. Kohlberg (1976) proposed that children develop a sense of what is right versus wrong, but that this sense is initially based on subjective, ego-centered feelings—whether the situation is self-relevant. More recently however, a growing body of work demonstrates that humans begin to have moral awareness as early as infancy (Hamlin, Wynn, & Bloom, 2007; 2010; Hamlin, Wynn, Bloom, & Mahajan, 2011; Kuhlmeier, Wynn, and Bloom, 2003). For example, Hamlin and colleagues demonstrated that at six months of age, infants will prefer a character who helped rather than hindered another character's goal. Schmidt and Sommerville (2011) found that 15-month-olds are sensitive to fairness and will spontaneously engage in altruistic behavior, demonstrating that infants have a basic sense of fairness and altruism (see also Sommerville, chap. 20, this volume; Sommerville, Schmidt, Yun, & Burns, 2013). Related research demonstrates that toddlers engage in other-regarding behavior as soon as they are physically able, such as reparative actions, showing guilt and empathy as evidenced by apologizing and responding to others'

suffering (Martin & Clark, 1982), and spontaneously engage in helping behavior (Martin & Olson, 2013).

Clearly, young children have an awareness of and will engage in actions that resemble moral behavior, but how does an emerging morality translate into the development of long-term moral character? Eisenberg and her colleagues conducted a series of longitudinal studies that examined prosociality and empathy in children and tracked these individuals over 20 years into early adulthood (see Eisenberg, Guthrie, Cumberland, Murphy, Shepard, Zhou et al., 2002; Eisenberg, Guthrie, Murphy, Shepard, Cumberland, & Carlo, 1999; Eisenberg-Berg & Hand, 1979; Eisenberg, Miller, Shell, McNalley, & Shea, 1991). Several key findings emerged from these studies that inform how moral character develops over time. First, high levels of empathy in preschoolers were associated with prosocial behavior across multiple contexts at preschool age (Eisenberg et al., 1999; Eisenberg & Fabes, 1998) and predicted situation-specific prosocial behaviors into late childhood and adolescence. These behaviors included donating of money (at ages 9–11), mothers' reports of the child's helpfulness (ages 15–16), and costly helping (assisting the experimenter for no extra pay; ages 17–18). A second key finding was that individual differences in overall self-reported prosociality over a variety of measures were remarkably stable from early adolescence into adulthood (Eisenberg et al., 2002). Third, Eisenberg's work demonstrated that empathy was positively associated with higher moral reasoning and higher social competence in children, and predicted higher scoring on these domains into adulthood (Eisenberg, 2005; Eisenberg et al., 1991; Murphy, Shepard, Eisenberg, Fabes, & Guthrie, 1999). In contrast, lower levels of empathy in childhood were associated with antisocial behaviors and externalizing problems (Zhou et al., 2002).

In addition to person-level variables, the development of moral behavior depends on contextual variables, such as parenting and the home environment. Prosocial behavior in early childhood has been shown to be associated with a parent being high in sympathy, allowing the child to express negative emotions openly, low levels of hostile emotions at home, and effective parenting techniques (Eisenberg, 2005). Additionally, parental warmth and expressivity have an impact on the child's empathy and social functioning (Zhou et al., 2002). These early environmental factors are critical to moral development over time as well: child-rearing practices used with five-year-olds have been shown to predict children's overall empathy years later, into their thirties (see Koestner, Franz, & Weinberger, 1990).

Given this research, clearly both environmental and dispositional factors are crucial contributors to the development of moral character. However, it remains largely unknown how these two factors interact to jointly influence moral development in youth. Future research will need to examine the reciprocal relationship between dispositional tendencies and situational and environmental influences to better understand how moral character can be firmly

developed in young children, and how to promote moral behavior in socially difficult situations.

Individual Differences in Moral Character Become Apparent in Adolescence and Adulthood

Blasi (1980, 1983) proposed that the extent to which individuals possesses moral character will critically depend on whether they deem moral qualities to be important to their self-understanding. Adolescence, therefore, is a critical time for the emergence of moral character, as the self becomes more firmly constructed (Erikson, 1968; Hart & Carlo, 2005), and as youth make simultaneous improvements in the skills that are relevant to moral behavior, such as perspective taking, self-understanding, and social problem-solving (Eisenberg, Cumberland, Guthrie, Murphy, & Shepard, 2005). With these person-level changes come simultaneous changes in the adolescents' environment, including new responsibilities, increased expectations, and the introduction of romantic relationships (Pardini, Loeber, & Stouthamer-Loeber, 2005). Because of these changes in person and environmental variables, individual differences in moral character are thought to become more apparent in adolescence and adulthood (Hart & Carlo, 2005).

Research suggests that there are important dispositional and situational differences between adolescents that develop strong moral character and those that do not. Matsuba and Walker (2004) administered a five-factor personality inventory (NEO-PI-R, Costa, & McCrae, 1992), an attachment scale, and interview measures of moral reasoning and faith development to adolescent-aged moral exemplars and comparison individuals matched for age, gender, educational background, and ethnicity. Exemplars were higher in agreeableness, especially for the facets of straightforwardness, modesty, and empathy. Beyond overall trait differences, exemplars showed better reasoning skills on moral dilemmas, were more willing to enter into close relationships, and were higher in faith development.

Related research has found that adolescent moral exemplars are more likely to emphasize academic achievement (Hart & Fegley, 1995) and are more committed to helping those in need (Lawford, Pratt, Hunsberger, & Pancer, 2005). In contrast, adolescents low in empathy have been shown to have a higher involvement in immoral behaviors, such as bullying and aggression (Gini, Albiero, Benelli, & Altoe, 2007). Individual differences in adolescent moral character are also associated with important situational and contextual factors, particularly parenting practices. Eisenberg-Berg and Mussen (1978) found that maternal parenting practices that were non-punitive, egalitarian, and encouraged open discussion of problems were associated with higher levels of empathy, especially in boys.

Studies of moral character in adulthood have also found considerable individual differences between exemplars and comparison subjects. Walker and

Frimer (2007) examined two types of moral exemplars (classified as either brave or caring) compared to matched comparison subjects. Their findings revealed that exemplar individuals had higher agency and communion traits, optimism, sensitization to the needs of others, and a healthy, secure attachment style. Interestingly, the caring exemplars were more nurturant, optimistic, and generative than the brave exemplars. Other research on moral exemplars has found several qualities that are common across exemplars, including increased empathy, a long-standing commitment to moral values, a capacity for change, positivity, underlying faith, and internalized altruistic values (Colby & Damon, 1992; Midlarsky, Jones, & Corely, 2005).

Related to the concept of moral exemplars is the degree to which individuals are motivated and actively try to achieve moral excellence. For example, Noftle (see chapter 22) examined "character strivings" in young adults: their ratings of the character traits and moral self-attributes that they currently possess, and their ratings of their ideal traits and attributes that they strive to possess. Overall, young adults rated their ideal traits and attributes as higher than their current level, suggesting that individuals at this age are interested in and strive to be more moral than they currently are. These findings are encouraging, but more research will need to be done examining how and whether individuals can intentionally try to build moral character over time.

The Study of Development of Moral Character Requires an Analysis of Person and Situation Variables

Taken together, these findings suggest there are clear motivational, dispositional, and contextual factors that contribute to the development of moral character. Environmental factors, especially parenting, play a key role in the development of moral character in childhood and adolescence and can predict future moral behavior (such as prosociality and empathy). Dispositional factors play a strong role in development as well: possessing moral character traits early in life can predict healthy moral development later in life and social functioning in a variety of domains. Taking into account the clear influence of both situation and person variables on the development of moral character and that moral behavior appears to vary considerably across situations (see Aquino et al., 2009; Eisenberg, 2005), numerous researchers have recommended a more systematic analysis of the patterning of one's moral behavior across diverse situations—or one's moral style—in order to better understand moral development (see Lapsley & Narvaez, 2004; Penner, Escarraz, & Ellis, 1983). Given the influence of the motivational, situational, and dispositional factors, an examination of individuals' moral behavioral patterns has the potential to better reveal individual differences and to disentangle the specific situational and personality influences on the development of moral character.

Conclusion

We propose that research devoted to examining these five questions will be essential to advancing our knowledge and igniting the study of character. However, we emphasize that these questions are only a starting point for research on character. As has happened in the field of personality psychology, we predict that the study of character will experience a renaissance of its own; the diverse social-cognitive, trait, and biological studies of character we have presented in this chapter are evidence that this resurgence of interest has already begun. As research on character continues and progresses, there will be new and important questions for researchers to examine. Continued research on character holds the potential to advance our understanding of moral and deviant behavior, elucidate the relationship between normal personality and morality, predict moral development over time, and help us better understand how to promote moral behavior in society.[3]

References

Aquino, K., Freeman, D., Reed, A. II, Lim, V. K., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, *97* (1), 123.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258–290.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*, 150–166. doi:10.1177/1088868306294907

Benet, V., & Waller, N. G. (1995). The Big Seven factor model of personality description: Evidence for its cross-cultural generality in a Spanish sample. *Journal of Personality and Social Psychology*, *69*, 701–718.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, *88*(1), 1.

Blasi, A. (1983). Moral cognition and moral action: A theoretical perspective.*Developmental review*, *3*(2), 178–210.

[3] Some of the content and ideas presented in this chapter were borrowed from Jayawickreme, Meindl, Helzer, Furr, and Fleeson, 2014. This project was made possible through the support of the Templeton World Charity Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Wake Forest University or the Templeton World Charity Foundation. We would like to thank Christian Miller, Brandon Warmke, and Kassidy Velasquez for their extensive and thoughtful help in developing these ideas.

Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *European Journal of Personality*, *23*(6), 509–530. doi:10.1002/per.731

Block, J. (1971). *Lives through time*. Berkeley, CA: Bancroft Books

Cascio, C. N., Konrath, S., & Falk, E. B. (2014). Narcissists' social pain seen only in the brain. *Social cognitive and affective neuroscience*. doi:10.1093/scan/nsu072

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science*, *20*, 1033–1039. doi:10.1111/j.1467-9280.2009.02409.x

Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psycho-physiological source traits in a normal individual. *Psychometrika*, *12*, 267–288.

Clark, L. A., & Watson, D. (1999). Personality, disorder, and personality disorder: Towards a more rational conceptualization. *Journal of Personality Disorders*, *13*, 142–151.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2013). Agreement and similarity in self-other perceptions of moral character. *Journal of Research in Personality*, *47*(6), 816–830.

Colby, A., & Damon, W. (1992). *Some do care: Contemporary lives of moral commitment*. New York: Free Press.

Conway, J. M., Rogelberg, S. G., & Pitts, V. E. (2009). Workplace helping: Interactive effects of personality and momentary positive affect. *Human Performance*, *22*, 321-339.

Cooley, C. H. (1902). *Human nature and the social order*. New York: Scribner's.

Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, *38*, 668–678.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEOPI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Darley, J. M., & Batson, C. D. (1973). 'From Jerusalem to Jericho': A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*, 100–108.

DeYoung, C. G. (2010). Personality neuroscience and the biology of traits. *Social and Personality Psychology Compass*, *4*(12), 1165–1180.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, *93*, 880–896. doi:10.1037/0022-3514.93.5.880

Diener, E., & Larsen, R.J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology*, *47*, 871–883.

Doris, J. M. (1998). Persons, situations and virtue ethics. *Noûs*, *32*, 504–530.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.

Eisenberg, N. (2005). The development of empathy-related responding. *Nebraska Symposium on Motivation*, *51*(January), 73–117.

Eisenberg, N., Cumberland, A., Guthrie, I. K., Murphy, B. C., & Shepard, S. A. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *Journal of research on adolescence*, *15*(3), 235-260.

Eisenberg, N., & Fabes, R.A. (1998). Prosocial development. In W. Damon & N. Eisenberg (Eds.) *Handbook of child psychology: Vol. 3. Social, emotional, and personality development*, 5th ed., pp. 701-778. New York: Wiley.

Eisenberg, N., Guthrie, I. K., Cumberland, A., Murphy, B. C., Shepard, S. A., Zhou, Q., & Carlo, G. (2002). Prosocial development in early adulthood: A longitudinal study. *Journal of Personality and Social Psychology*, *82*(6), 993.

Eisenberg, N., Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development*, *70*(6), 1360–1372.

Eisenberg-Berg, N., & Hand, M. (1979). The relationship of preschoolers' reasoning about prosocial moral conflicts to prosocial behavior. *Child Development*, *50*, 356–363.

Eisenberg, N., Miller, P. A., Shell, R., McNalley, S., & Shea, C. (1991). Prosocial development in adolescence: A longitudinal study. *Developmental Psychology*, *27*(5), 849.

Eisenberg-Berg, N., & Mussen, P. (1978). Empathy and moral development in adolescence. *Developmental Psychology*, *14*(2), 185.

Erikson, E. H. (1968). *Identity: Youth and Crisis* (No. 7). New York:W. W. Norton.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*, 1097–1126. doi:10.1037/0022-3514.37.7.1097

Fleeson, W. (2001). Towards a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027. doi:10.1037/0022-3514.80.6.1011

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, *75*, 825–861.

Fleeson, W., Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The prospects for a personality-based perspective on morality. *Social and Personality Psychology Compass*, *8*(4), 178–191.

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*, 1097–1114. doi:10.1037/a0016786.

Fleeson, W., Malanos, A. B., & Achille, N.M. (2002). An intraindividual process approach to the relationship between extraversion and positive affect: Is acting extraverted as 'good' as being extraverted? *Journal of Personality and Social Psychology*, *83*, 1409–1422. doi:10.1037/0022-3514.83.6.1409

Fleeson, W., & Noftle, E. E. (2008). Where does personality have its influence? A supermatrix of consistency concepts. *Journal of Personality*, *76*(6), 1355–1386. doi:10.1111/j.1467-6494.2008.00525.x.

Fournier, M., Moskowitz, D., & Zuroff, D. (2009). The interpersonal signature. *Journal of Research in Personality*, *43(2),* 155–162. doi:10.1016/j.jrp.2009.01.023

Fridhandler, B. M. (1986). Conceptual note on state, trait, and the state-trait distinction. *Journal of Personality and Social Psychology*, *50*, 169–174.

Funder, D. C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality and Social Psychology*, *48*, 473–493.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, *101*(1), 75–90.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670. doi:10.1037/0033-295X.102.4.652

Furr, R.M., Dougherty, D. M., Marsh, D. M., & Mathias, C. W. (2007). Personality judgment and personality pathology: Self-other agreement in adolescents with conduct disorder. *Journal of Personality*, *75*(3), 629-662.

Gebauer, J. E., Riketta, M., Broemer, P., and Maio, G. R. (2008). "How much do you like your name?" An implicit measure of global self-esteem. *Journal of Experimental Social Psychology*, *44*(5), 1346–1354.

Gini, G., Albiero, P., Benelli, B., & Altoè, G. (2007). Does empathy predict adolescents' bullying and defending behavior? *Aggressive Behavior*, *33*(5), 467–476.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148.

Hall, J., Andrzejewski, S., Murphy, N., Mast, M., & Feinstein, B. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, *42*, 1476–1489. doi:10.1016/j.jrp.2008.06.013

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences*, *108*(50), 19931–19936.

Hampson, S. E., Edmonds, G. W., Goldberg, L. R., Dubanoski, J. P., & Hillier, T. A. (2013). Childhood conscientiousness relates to objectively measured adult physical health four decades later. *Health Psychology*, *32*(8), 925.

Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society*, *99*, 315–31.

Hartshorne, H. & May, M. A. (1928). *Studies in the Nature of Character, vol. I: Studies in Deceit.* New York: Macmillan.

Hart, D., & Carlo, G. (2005). Moral development in adolescence. *Journal of Research on Adolescence*, *15*(3), 223–233.

Hart, D., & Fegley, S. (1995). Prosocial behavior and caring in adolescence: Relations to self-understanding and social judgment. *Child Development*, *66*(5), 1346–1359.

Heller, D., Komar, J., & Lee, W. B. (2007). The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin*, *33*, 898–910.

Helzer, E. G., Furr, R. M., Hawkins, A., Barranti, M., Blackie, L. E., & Fleeson, W. (2014). Agreement on the perception of moral character. *Personality and Social Psychology Bulletin*, *40*, 1698-710. doi: 10.1177/0146167214554957

Ho, S. S., Konrath, S., Brown, S., & Swain, J. E. (2014). Empathy and stress-related neural responses in maternal decision-making. *Frontiers in Neuroscience*, *8*, 152. doi: 10.3389/fnins.2014.00152

Hopwood, C. J., Donnellan, M. B., Blonigen, D. M., Krueger, R. F., McGue, M., Iacono, W. G., & Burt, S. A. (2011). Genetic and environmental influences on personality trait

stability and growth during the transition to adulthood: A three-wave longitudinal study. *Journal of Personality and Social Psychology*, *100*(3), 545.

Hunt, E. (1965). Selection and reception conditions in grammar and concept learning. *Journal of Verbal Learning & Verbal Behavior*, *4*, 211–215.

Jayawickreme, E., Meindl, P., Helzer, E. G., Furr, R. M., & Fleeson, W. (2014). Virtuous states and virtuous traits: How the empirical evidence regarding the existence of broad traits saves virtue ethics from the situationist critique. *Theory and Research in Education*. doi:1477878514545206

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3*, 114-158.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24.

Kandler, C., Bleidorn, W., Riemann, R., Spinath, F. M., Thiel, W., & Angleitner, A. (2010). Sources of cumulative continuity in personality: A longitudinal multiple-rater twin study. *Journal of Personality and Social Psychology*, *98*(6), 995.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*, 23–34.

Kern, M. L., Duckworth, A. L., Urzúa, S. S., Loeber, R., Stouthamer-Loeber, M., & Lynam, D. R. (2013). Do as you're told! Facets of agreeableness and early adult outcomes for inner-city boys. *Journal of Research in Personality*, *47*(6), 795–799.

Koestner, R., Franz, C., & Weinberger, J. (1990). The family origins of empathic concern: A 26-year longitudinal study. *Journal of Personality and Social Psychology*, *58*(4), 709.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive–developmental approach. In T.Lickona (Ed.), *Moral development and behavior* (pp. 31– 53). New York: Holt, Rinehart & Winston.

Krueger, R., South, S., Johnson, W., & Iacono, W. (2008). The heritability of personality is not always 50%: Gene-environment interactions and correlations between personality and parenting. *Journal of Personality*, *76*, 1485–1522. doi:10.1111/j.1467-6494.2008.00529.x

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402-408.

Lapsley, D. K., & Narvaez, D. (2004). A social-cognitive approach to the moral personality. In D. K. Lapsley & D. Narvaez (Eds.) *Moral development, self, and identity* (pp. 189–212). Mahwah, NJ: Lawrence Erlbaum.

Latané, B., & Darley, J. M. (1970). *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton-Century Crofts.

Lawford, H., Pratt, M. W., Hunsberger, B., & Mark Pancer, S. (2005). Adolescent generativity: A longitudinal study of two possible contexts for learning concern for future generations. *Journal of Research on Adolescence*, *15*(3), 261–273.

Lucas, R. E., & Fujita, F. (2000). Factors influencing the relation between extraversion and pleasant affect. *Journal of Personality and Social Psychology*, *79*, 1039–1056.

Martin, G. B., & Clark, R. D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology*, *18*(1), 3.

Martin, A., & Olson, K. R. (2013). When kids know better: Paternalistic helping in 3-year-old children. *Developmental Psychology*, *49*(11), 2071.

Matsuba, M. K., & Walker, L. J. (2004). Extraordinary moral commitment: Young adults involved in social organizations. *Journal of Personality*, *72*(2), 413–436.

McAdams, D. P., & Olson, B. D. (2010). Personality development: Continuity and change over the life course. *Annual Review of Psychology*, *61*, 517–542.

McNiel, J. M., Lowman, J. C., & Fleeson, W. (2010). The effects of state extraversion on four different types of affect. *European Journal of Personality*, *24*, 18–35.

Mead, G. (1934). *Mind, self, and society from the perspective of a social behaviorist*. Chicago: University of Chicago Press.

Midlarsky, E., Jones, S., & Corley, R. P. (2005). Personality correlates of heroic rescue during the Holocaust. *Journal of Personality*, *73*(4), 907–934.

Miller, C. B. (2013). *Moral Character: An Empirical Theory*. New York: Oxford University Press.

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*, 252–283.

Moskowitz, D. S. (1982). Coherence and cross-situational generality in personality: A new analysis of old problems. *Journal of Personality and Social Psychology*, *43*, 754–768.

Murphy, B. C., Eisenberg, N., Fabes, R. A., Shepard, S., & Guthrie, I. K. (1999). Consistency and change in children's emotionality and regulation: A longitudinal study. *Merrill-Palmer Quarterly (1982)*, 413-444.

Narvaez, D., Lapsley, D. K., Hagele, S., & Lasky, B. (2006). Moral chronicity and social information processing: Tests of a social cognitive approach to the moral personality. *Journal of Research in Personality*, *40*(6), 966–985.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Noftle, E. E., & Robins, R. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*, 116–130. doi:10.1037/0022-3514.93.1.116

Ones, D., Viswesvaran, C., & Schmidt, F. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, *78*, 679–703. doi:10.1037/0021-9010.78.4.679

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421.

Pardini, D. A., Loeber, R., & Stouthamer-Loeber, M. (2005). Developmental shifts in parent and peer influences on boys' beliefs about delinquent behavior.*Journal of Research on Adolescence*, *15*(3), 299-323.

Penner, L. A., Escarraz, J., & Ellis, B. B. (1983). Sociopathy and helping: Looking out for number one. *Academic Psychology Bulletin*, *5*, 195–209.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126(1)*, 3–25.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L.R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313–345.

Roberts, B. W., & Wood, D. (2006). Personality development in the context of the neo-socioanalytic model of personality. In Mroczek, Daniel K., and Little, Todd D. (Eds.) *Handbook of Personality Development* (pp. 11–39). Mahwah, NJ: Lawrence Erlbaum.

Romero-Canyas, R., Anderson, V. T., Reddy, K. S., & Downey, G. (2009). Rejection sensitivity. In M. R. Leary, R. H. Hoyle, M. R. Leary, & R. H. Hoyle (Eds.), *Handbook of Individual Differences in Social Behavior* (pp. 466–479). New York: Guilford Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, *10*, 173–220.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.

Schmidt, M. F., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS One*, *6*(10), e23223.

Shiner, R. L., Masten, A. S., & Roberts, J. M. (2003). Childhood personality foreshadows adult personality and life outcomes two decades later. *Journal of Personality*, *71*(6), 1145–1170.

Shrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin*, *86(3),* 549-573.

Sommerville, J. A., Schmidt, M. F., Yun, J. E., & Burns, M. (2013). The development of fairness expectations and prosocial behavior in the second year of life. *Infancy*, *18*(1), 40–66.

Stautz, K., & Cooper, A. (2013). Impulsivity-related personality traits and adolescent alcohol use: A meta-analytic review. *Clinical Psychology Review*, *33*(4), 574–592.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171.

Swann, W. B., Jr., & Seyle, C. (2005). Personality psychology's comeback and its emerging symbiosis with social psychology. *Personality and Social Psychology Bulletin*, *31*(2), 155–165.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*, 160–164.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98(2),* 281–300.

Vernon, P. E. (1964). *Personality assessment: A critical survey*. New York: Wiley.

Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, *93*(5), 845.

Wallace, J. (1966). An abilities conception of personality: Some implications for personality measurement. *American Psychologist*, *21*, 132–138.

Zhou, Q., Eisenberg, N., Losoya, S. H., Fabes, R. A., Reiser, M., Guthrie, I. K., & Shepard, S. A. (2002). The relations of parental warmth and positive expressiveness to children's empathy-related responding and social functioning: A longitudinal study. *Child Development*, *73*(3), 893–915.

PART } II

Beliefs about Character

3 }

Lay Beliefs in True Altruism versus Universal Egoism

Jochen E. Gebauer, Constantine Sedikides,
Mark R. Leary, and Jens B. Asendorpf

Contemplating the first and most simple operations of the human soul,
I think I can perceive in it two principles prior to reason,
one of them deeply interesting us in our own welfare and preservation,
and the other exciting a natural repugnance at seeing any other sensible
being, and particularly any of our own species, suffer pain or death.

—ROUSSEAU, *1950/1755, PREFACE*

No man giveth but with intention of good to himself,
because gift is voluntary; and of all voluntary acts,
the object is to every man his own good; of which,
if men see they shall be frustrated, there will be no beginning of
benevolence or trust, nor consequently of mutual help.

—HOBBES, *1651, CHAPTER 15*

Are human beings capable of truly altruistic behavior? This question has been central to many scholarly disciplines, including philosophy, psychology, economics, and theology (Batson, 1991, 2011). Yet scholars still disagree fervently on the answer. We approached this issue indirectly. Our research did not concern the existence of true altruism or scholarly disagreement on the topic. Instead, it focused on the beliefs of lay people.

Do lay people side with Rousseau's introductory claim and deem that people are capable of truly altruistic behavior? Alternatively, do they side with Hobbes's claim and agree that all apparent altruism is egoistically motivated? Moreover, is there consensus or disagreement among lay people's beliefs? If lay people differed in their beliefs in true altruism, how would these differences be linked to their other beliefs about human nature? Finally, and probably most importantly, are individual differences regarding belief in the existence of true altruism associated with moral character? These questions

have potentially important ramifications, because the answers promise a fuller understanding of the genesis of moral character. Thus, our research (Gebauer, Sedikides, Leary, & Asendorpf, 2015), which we review in this chapter, has begun to address these questions.

Our research may also help to bridge the classic altruism literature (Batson, 2011; Sober & Wilson, 1998; Stich, Doris, & Roedder, 2010) with the burgeoning literature on lay beliefs (Baumeister, 2008; Carey & Paulhus, 2013; Dweck, 2000). Hence, before embarking on a description of our findings, we briefly discuss (a) the literature on true altruism, (b) the broader literature on lay beliefs, and (c) prior indirect evidence regarding lay beliefs in true altruism versus universal egoism.

Brief Review of the True Altruism Literature

Helpfulness, kindness, cooperation, and the like, are frequent human behaviors (Dovidio, Piliavin, Schroeder, & Penner, 2006). There is no scientific dispute about that. There is also no scientific dispute about the notion that these prosocial behaviors can be motivated by various forms of self-interest. Put otherwise, even the most fervent proponent of true altruism's existence would agree that prosocial behavior is often motivated by egoistic concerns, including the expectation of personal benefit, positive affect, or reciprocal favor (Batson, 2011; Rousseau, 1950/1755). Hence, the altruism debate does not revolve around the issues of whether people act prosocially (they do) or whether human prosociality can be egoistically motivated (it can). The altruism debate refers to whether human prosociality is sometimes motivated by a genuine desire to benefit others as an end in itself—that is, by true altruism.

We borrowed our definition of true altruism from Batson (1991, p. 6): "Altruism is a motivational state with the ultimate goal of increasing another's welfare." In contrast to Batson, we use the term *true altruism* rather than simply *altruism*. We were influenced in our terminological choice by researchers who suggested that the term *true altruism* better describes Batson's motivational conceptualization of altruism (Maner, Luce, Neuberg, Cialdini, Brown, & Sagarin, 2002; Piliavin & Charng, 1990; Sober & Wilson, 1998), because the simple term *altruism* has frequently served as a synonym for reciprocity (Trivers, 1971) and prosociality (Rushton, 1976; Sober, 1988).

Empirical Approaches to the Existence of True Altruism

Traditionally, the altruism question has been mainly debated in philosophy as well as in theology and economics (for thorough reviews, see Batson, 2011; Stich et al., 2010). Yet, over the last half-century, psychology has been at the

forefront of the true altruism debate. Even philosophers welcome this shift: "As generations of introductory philosophy students have found out, such arguments [i.e., logical arguments for true altruism versus universal egoism] are bound to end in inconclusive speculation—so long as the competing explanations are not *empirically* evaluated" (Stich et al., 2010, p. 170; emphasis added). What has psychological research found out?

Many of the early empirical psychologists assumed that human nature is characterized by universal egoism (Campbell, 1975). Guided by this assumption, many psychologists in the 1970s amassed evidence for egoistic reasons underlying helping (Cialdini, Darby, & Vincent, 1973; Piliavin & Piliavin, 1973; Schwartz, 1977; Staub, 1974). For example, Schwartz (1977) pointed to the power of normative expectations as drivers of prosociality. Some others, however, made an empirical case for true altruism (Krebs, 1975; Meindl & Lerner, 1983; Sagi & Hoffman, 1976). This case, though, was inconclusive. As Batson (1991) put it, "[T]he empirical evidence presented in support is unconvincing; it can easily be accounted for by one or more egoistic explanations" (p. 47).

Still, the early empirical evidence for true altruism was crucial, because it highlighted the role of other-oriented emotions as motivational forces of true altruism. For example, Hoffman (1981) saw "an affective response appropriate to someone else's situation rather than one's own" (p. 128) as the motivational basis of true altruism. Similarly, Rosenhan, Salovey, Karylowski, and Hargis (1981) argued that "while experiencing those [other-oriented] affects, people are led away from stances that maximize their own rewards and find instead that their attention, cognitions, and behaviors are directed toward the needs of others, often without regard for the quid pro quo" (p. 234).

To be sure, the idea that other-oriented emotions are involved in the genesis of true altruism did not originate in psychology. Theologians and philosophers had entertained this possibility for centuries (Stich et al., 2010). For example, Thomas Aquinas (1917/1270) argued that "mercy is the heartfelt sympathy for another's distress, impelling us to succour him if we can" (article 1, objection 3), and Adam Smith (1853/1759) argued that "*compassion*, the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner" (section 1, chapter 1; emphasis added) drives selfless concern for others. Nonetheless, early empirical psychologists introduced other-oriented emotions into the empirical inquiry of true altruism and, as such, they paved the way for the most influential statement of true altruism, Batson's (2011) theory.

At the heart of Batson's theory lies the empathy-altruism hypothesis (Batson, 1987, 1991), according to which "empathic concern produces altruistic motivation" (Batson, 2011, p. 11). We have described how Batson defines altruism (i.e., *true altruism* in our terminology). But what precisely does Batson mean by the term "empathy"? Empathy, he stated, refers to "other-oriented emotion elicited by and congruent with the perceived welfare of someone in need" (Batson, 2011, p. 11). He clarified that empathy is ultimately "other-oriented in

the sense that it involves feeling for the other—feeling sympathy for, compassion for, sorry for, distressed for, concerned for, and so on" (Batson, 2011, p. 11). The empathy-altruism hypothesis stipulates that perceiving a person in need, coupled with adopting that person's perspective, can induce empathic concern for the other person. This, in turn, can trigger the truly altruistic motivation to have the other person's need satisfied.

Batson and colleagues first presented empirical evidence for the empathy-altruism hypothesis in the late 1970s and early 1980s (Batson, Duncan, Ackerman, Buckley, & Birch, 1981; Coke, Batson, & McDavis, 1978). In subsequent years, skeptics of true altruism's existence raised several egoistic alternatives underlying the effect of empathy on prosociality. These alternatives include aversive-arousal reduction (Piliavin, Dovidio, Gaertner & Clark, 1981), empathy-specific punishment (Archer, 1984; Dovidio, 1984), general empathy-specific reward (Thompson, Cowan, & Rosenhan, 1980), empathic joy as empathy-specific reward (Smith, Keating, & Stotland, 1989), negative-state relief as empathy-specific reward (Cialdini, Schaller, Houlihan, Arps, Fulz, & Beaman, 1987), and empathy-specific self-other merging (Cialdini, Brown, Lewis, Luce, & Neuberg, 1997; for reviews, see Batson, 1991, 2011; Batson & Powell, 2003; Cialdini, 1991).

Today, after 40 years of empirical scrutiny, how does the empathy-altruism hypothesis fare? Contemporary psychologists' views remain divided, partly because definitive empirical tests have been difficult to obtain. To illustrate, Batson and Powell (2003) concluded that contemporary alternatives cannot explain the evidence for the empathy-altruism hypothesis; hence "in addition to our all-too-apparent failing and fallibilities, we humans are, at times, capable of caring, and caring deeply, for people and issues other than ourselves" (p. 479). In direct opposition to this conclusion, Maner et al. (2002, p. 1609) concluded that "evidence gathered across numerous studies compellingly points to egoistic, rather than altruistic, explanations for prosocial behavior, including those forms of aid for which purely selfless motives have been claimed."

The long history of belief versus disbelief in true altruism's existence among scientists is interesting in its own right. This, however, is not the reason we described the state of scientific disagreement at some length. The reason is that this long-standing disagreement can also be seen in lay people's beliefs about altruism.

Brief Review of the Literature on Lay Beliefs

Lay beliefs can be impactful guides of behavior (Baumeister, Masicampo, & DeWall, 2009), comparable perhaps to attitudes (Ajzen & Fishbein, 1980), values (Rokeach, 1973), goals (Kruglanski, Shah, Fishbach, Friedman, Chun, &

Sleeth-Keppler, 2002), and personality traits (John & Srivastava, 1999). Yet, the literature on lay beliefs is considerably less developed and influential than the literatures on the aforementioned constructs. Relative lack of influence may be due to its heterogeneity and disconnect from the mainstream.

In this section, we briefly review the research on lay beliefs. We discuss beliefs in a just world, in free will, in the existence of God, and in moral goodness. Further, we highlight the relation between these lay beliefs and moral character.

FOUR LAY BELIEFS AND THEIR RELATION TO MORAL CHARACTER

Belief in a just world. According to Lerner and Miller (1978), people have a "need to believe that they live in a world where people generally get what they deserve" (p. 1030). As a result of this need, lay people are prone to believing that they live in a just world (Lerner, 1965, 1980). However, people differ in the degree to which they consider the world just (Dalbert, Montada, & Schmitt, 1987; Rubin & Peplau, 1975). Such individual differences are related to life outcomes (Dalbert, 2009; Furnham, 2003), including moral character. Specifically, belief in a just world can motivate prosocial behavior, because just-world believers may assume that acting prosocially will be rewarded, whereas refraining from prosocial acts will be penalized (Bierhoff, Klein, & Kramp, 1991).

Belief in free will. The existence of free will has sparked long and irreconcilable debates in philosophy, much as the existence of true altruism has (Baumeister, Mele, & Vohs, 2010; Kane, 2005). Unlike the topic of lay belief in the existence of true altruism, lay belief in free will has garnered a great deal of psychological attention over the past 30 years (Baumeister, 2008; Stillman, Baumeister, Vohs, Lambert, Fincham, & Brewer, 2010; Stroessner & Green, 1990; Viney, Waldman, & Barchilon, 1982). Relevant research has documented that most lay people believe in the existence of free will (Paulhus & Carey, 2011).

This research has also identified individual differences in the belief in free will and shown that these individual differences are associated with life outcomes, including moral character. Free will believers (compared to disbelievers) exhibit more helping behavior (Baumeister et al., 2009). Indeed, experimental evidence buttresses the idea that belief in free will encourages moral behavior; whereas disbelief in free will discourages moral behavior. For example, persuading people that free will does not exist leads to higher levels of cheating (Vohs & Schooler, 2008), lower willingness to help, and increased aggression (Baumeister et al., 2009).

Individual differences concerning belief in free will are typically assessed with continuous rating scales, ranging from a strong conviction that free will does not exist to a strong conviction that it does (Baumeister et al., 2009; Carey & Paulhus, 2013; Vohs & Schooler, 2008). The advantage of continuous rating

scales is that they allow participants to express differing degrees of subjective certainty about the existence (or non-existence) of free will. Our research on belief in true altruism adopted this continuous rating scale approach (Gebauer et al., 2015). Our use of continuous rating scales is also consistent with research on other lay beliefs, including the belief that personality is mutable versus fixed (Chiu, Hong, & Dweck, 1997; Dweck, 2000; Dweck & Leggett, 1988) and the belief in the existence versus non-existence of God (Friese & Wänke, 2014; Kay, Gaucher, Napier, Callan, & Laurin, 2008; Shariff & Norenzayan, 2007).

Belief in the existence of God. Belief in the existence of God is one of the oldest documented lay beliefs of humankind. The Venus of Willendorf is a 4.25-inch statuette, probably carved in the Paleolithic Period ("Old Stone Age"). The statuette is thought to symbolize a fertility goddess (Ruether, 2005) and, as such, dates belief in god(s) at least to 22,000 BC. Philosophical debates about the existence of God date back to the ancient Greeks (see Aristotle's *Metaphysics*), much like debates about the existence of true altruism and free will. Compared to the latter two kinds of lay belief, however, belief in God's existence has received much empirical attention from psychologists (Baumeister, 2002; Saroglou, 2014; Sedikides, 2010).

Belief in the existence of God is prevalent. A representative survey from the 1980s showed that 95% of the US population believed in God's existence (Gallup & Castelli, 1989). Nonetheless, there are individual differences linked to various outcomes. For example, religious belief (including belief in God) is associated with higher self-control (McCullough & Willoughby, 2009). Although one might expect that religious people would be more prosocial (Shariff & Norenzayan, 2007), the relationship between belief in God and prosociality has proven to be intricate. Religious individuals report considerably more prosocial behavior (Bloom, 2012; Galen, 2012), and informants attribute heightened prosociality to them (Saroglou, 2012; Saroglou, Pichon, Trompette, Verschueren, & Dernelle, 2005). The picture changes, however, when prosociality is operationalized as overt behavior (Darley & Batson, 1973; Norenzayan & Shariff, 2008). One extensive review of that literature concludes that the religious prosociality hypothesis is altogether a "congruence fallacy" (Galen, 2012, p. 899). The *minimal prosociality* account (Saroglou, 2013) offers another explanation for the gap between religious individuals' prosociality self-reports and their behavior. According to Saroglou (2013), religious individuals exhibit rather specific forms of prosocial behavior, "behaviors that are not necessarily of high cost but hold at least some importance if one wants to perceive oneself and be perceived by others as moral" (p. 4).

Belief in moral goodness. Are people generally good to others or are people generally bad and defective in cooperative situations, refuse to help, and lack trustworthiness? Lay belief in moral goodness has been studied by many psychologists under various labels. For example, Rosenberg's (1956) Faith in People scale assesses belief in moral goodness with items such as "Most people

are inclined to help others" and "If you don't watch yourself, people will take advantage of you" (negatively scored). Schuessler's (1982) Doubt about Trustworthiness of People scale assesses belief in moral goodness with items such as "Most people can be trusted" and "Most people don't really care what happens to the next fellow" (negatively scored). Wrightsman (1991) noted the conceptual overlap between these measures. Content validity considerations suggest that these measures are best described as indicators of the belief that human beings generally possess a morally good character.

There is a key difference between belief in moral goodness and beliefs in a just world, free will, and God's existence. Specifically, most lay people do not believe that humans are generally good-natured and possess moral character. Participants from France, Germany, and the United States provided ratings that are consistently closer to the "bad nature" than the "good nature" pole of the scale (Bégue, 2002; Gebauer et al., 2015). However, here again, individual differences emerge (Rosenberg, 1956; Schuessler, 1982). Also, lay people who believe that people are generally good-natured possess more moral character (Cadenhead & Richman, 1996; Rosenberg, 1957). Experimental evidence documents the effect of belief in moral goodness on prosocial behavior. Hornstein, LaKind, Frankel, and Manne (1975) exposed participants to news broadcasts that highlighted humans' good nature or their bad nature. Participants who viewed the good nature broadcast exhibited more prosocial behavior than those who viewed the bad nature broadcast.

There is a conceptual difference between belief in moral goodness and belief in the existence of true altruism. Belief in moral goodness refers to whether people on average behave prosocially and morally (Wrightsman, 1991). In contrast, belief in the existence of true altruism refers to the *motivation* that underlies such prosocial and moral behavior. Hence, a person who believes that people are generally prosocial and moral may also believe that their prosociality and morality are based on true altruism; alternatively, this person may believe that all prosociality and morality are based on universal egoism.

CONCLUSIONS ABOUT LAY BELIEFS

In all, empirical research suggests that most lay people believe in a just world, free will, and the existence of God. At the same time, this research suggests that most lay people have doubts about people's moral goodness and believe that human behavior is more bad than good. Individual differences in these lay beliefs have been linked to life outcomes, in general, and to prosociality and moral character, in particular. In contrast, there is little, if any, conclusive research on lay beliefs in the existence of true altruism. Still, existing research provides a starting point for our own empirical endeavors (Gebauer et al., 2015). We turn to that prior research next.

Brief Review of Prior Research on Lay Beliefs in True Altruism versus Universal Egoism

Two research lines pertain to individual differences in belief in the existence of true altruism. In their early work on altruism, Batson and his colleagues used several self-report measures to assess "self-perceived altruism" (Batson, Coke, Jasnoski, & Hanson, 1978; Batson, Fulz, Schoenrade, & Paduano, 1987; Thomas & Batson, 1981; Thomas, Batson, & Coke, 1981). With one exception, that work assessed self-perceived altruism with self-ascription of multiple traits, including "helpful," "cooperative," "compassionate," "considerate," "unselfish," "generous," "responsible," and "sympathetic." These measures, then, assess neither participants' belief that they act truly altruistically nor their belief that humans in general are capable of acting in a truly altruistic manner. Instead, the measures are better described as assessing a broad, prosociality-related personality factor, which is often labeled "communion" (Gebauer, Leary, & Neberich, 2012). Despite the label "self-perceived altruism," then, that research does not bear relevance to lay beliefs in the existence of true altruism, and it did not intend to do so.

The exception is Batson et al.'s (1987) Experiment 1. Self-perceived altruism was assessed with two items: "To what extent do you generally help others for selfish reasons?" and "To what extent do you generally help others for unselfish reasons?" (1 = *not at all*, 9 = *extremely*). These items assess participants' belief that they act truly altruistically and thus provide indirect insight into lay beliefs in the existence of true altruism; after all, one must believe in true altruism's existence in general in order to belief that one's own actions are at times driven by true altruism. Batson et al.'s results, then, are informative for our research questions. These results revealed that participants believed they generally help others more for unselfish reasons ($M = 5.39$) than for selfish reasons ($M = 2.77$)—a statistically significant difference. Thus, lay people believe not only that true altruism exists but also that most of their prosocial behavior is motivated by unselfish reasons. Yet, the results should be interpreted with caution. It is unclear whether participants' implicit definition of selfish versus unselfish reasons to help is consistent with the definition of true altruism that we have set forth here (see also Batson, 1991, 2011).

The second relevant line of research involves work on the belief in pure evil and pure good (Webster & Saucier, 2013). The 28-item Belief in Pure Good scale includes five items that assess belief in the existence of true altruism (e.g., "There is such a thing as a truly selfless/altruistic person"; "People only help others because they expect to be rewarded" [reverse scored]). However, although a handful of other scale items are indirectly linked to true altruism (e.g., "In essence, 'pure good' is selflessly helping other people in need"; "Even selfless people hate helping enemies" [reverse scored]), the vast majority of items are irrelevant to the existence of true altruism (e.g., "The forces of good will always prevail in the end"; "We DO NOT need more 'pure good' people in this world").

Unfortunately, Webster and Saucier reported results only for the full scale rather than for the items relevant to belief in the existence of true altruism separately. Hence, their results do not directly inform our research question. Nonetheless, it is telling that these authors found that true altruism items are part of the belief that humans can be purely good. This empirical finding is consistent with the entrenched philosophical position that the capacity for true altruism is a testimony to humans' good nature (Aristotle, 1911/350 BC; Bentham, 1876/1789). Two additional findings by Webster and Saucier are relevant. First, mean values of the Belief in Pure Good scale suggest that most people are inclined to believe in pure goodness: means were consistently above the theoretical scale midpoint. Second, the belief in pure goodness was related to some indicators of moral character, including self-reported volunteerism and support for needy children.

Together, a few studies have been concerned with the lay belief in the existence of true altruism, but they do not inform the questions that we posed (Gebauer et al., 2015). Specifically, no prior research has directly examined (a) whether lay people differ on their belief in the existence of true altruism and, if so, how many are "naive Rousseauians" versus "naive Hobbesians"; (b) whether and how belief in true altruism's existence relates to lay beliefs in a just world, free will, the existence of God, and moral goodness; and (c) whether and how belief in the existence of true altruism relates to prosociality and moral character. In the following three sections, we review our preliminary findings.

Review on Research Question I: Naive Rousseauians or Naive Hobbesians?

Are most lay people naive Rousseauians, believing that human prosociality can be truly altruistic, or naive Hobbesians, believing that human prosociality is universally egoistic? In a first step toward resolving these ambiguities, we (Gebauer et al., 2015) constructed a measure of belief in true altruism's existence, the Belief in the Existence of True Altruism (BETA) scale, which we describe below.

DEVELOPMENT OF THE BETA SCALE

Scale instructions. To be certain that participants understood the constructs of interest, we provided them with clear explanations of the key terms, including "selfless," "altruistic," "selfish," and "egoistic." Participants read the following instructions before responding to the items:

> Are humans capable of doing selfless good deeds or is all human behavior in the final end selfish (that is, motivated by expecting personal benefits, good feelings, social acceptance, or to be treated well in return)?

> In other words, can people be ultimately other-loving (i.e., be truly "altruistic") or does self-love (i.e., "egoism") underlie all human behavior? We are interested in your personal opinion. The following items will help us to assess whether you personally believe that people are capable of doing selfless good deeds or whether all human behavior is selfish in the final end (that is, motivated by expecting personal benefits, good feelings, social acceptance, or to be treated well in return).

Scale items. The 16 items of the initial BETA scale followed. Each item was accompanied by a 7-point rating scale (1 = *absolutely wrong*, 7 = *absolutely right*). Sample items are "People can act with the ultimate goal of contributing to another person's welfare and without expecting personal benefits of any kind"; "We have the potential to love other people in a truly altruistic and unselfish manner without expecting emotional, social, or material benefits for ourselves"; "Everything humans do for other people is motivated by self-interest of one kind or another" (negatively scored); and "The ultimate goal underlying our love for other people is selfish and egoistic in nature" (negatively scored).

Exploratory and confirmatory factor analyses. We included the 16-item BETA scale in two online studies. We advertised Study 1 ($N = 149$) on a US-based web portal that lists psychological online studies (in English). We advertised Study 2 ($N = 226$) on a German web portal for psychological online studies (in German). We used data from both studies to examine the factor structure of the 16 BETA items. The results of exploratory factor analyses (EFAs) were highly consistent across the studies. The EFA indicated the existence of two correlated dimensions (or factors). One factor contained the eight positively scored items; the other, the eight negatively scored items. An explanation for the two-factor structure is that lay people—unlike scholars (Batson, 2011)—do not think of true altruism and universal egoism as two opposing poles of a single dimension. Another explanation is to consider item scoring direction as an undesirable method factor that contaminated the single-factor structure of BETA. Indeed, scoring direction can artificially make a one-dimensional construct appear two-dimensional (Paulhus & Vazire, 2007), with the Rosenberg Self-Esteem Scale (Rosenberg, 1965) being a well-known example (Marsh, 1996; Tomas & Oliver, 1999).

We then used confirmatory factor analyses (CFAs) to examine whether the BETA scale's two-factor structure reflects a substantive, conceptual difference in positive versus negative beliefs about true altruism or a methodological response tendency to respond differently to positively worded versus negatively worded items. To do this, we conducted two additional online studies. We advertised Study 3 ($N = 171$) in the United States and Study 4 ($N = 183$) in Germany (in English and German, respectively). Among other measures, participants completed the BETA Scale, which we shortened to 10 items (5 positively scored, 5 negatively scored) on the basis of the two EFAs in Studies

1–2 . Results in both studies indicated that the fit of the one-factor model was superior to that of the two-factor model.

DESCRIPTIVE STATISTICS OF THE BETA SCALE

We proceeded to examine mean scores on the BETA scale, the distribution of test scores, and the relations of BETA with demographic variables (culture, age, sex). We pooled the data of Studies 1–4 to achieve a larger sample ($N = 729$).

Overall tendencies. We asked whether people on average tend to believe in the existence of true altruism versus universal egoism. Hence, we tested whether participants' test score on the BETA Scale was on average located above or below the theoretical scale midpoint (1 = *absolutely wrong*, 7 = *absolutely right*): scores above 4 indicated belief in true altruism; scores below 4, belief in universal egoism. Results revealed that on average participants tended to believe in the existence of true altruism. Both the mean and median of test scores across the pooled four-study sample were about one point above the theoretical midpoint of the scale. Conceptually identical results emerged when we separately examined the five positively scored items and the five negatively scored items. Lay people, then, are generally inclined to believe that true altruism exists.

Individual differences. We inspected the distribution of BETA scores. Do lay people uniformly believe in the existence of true altruism, or are there sizeable individual differences? If such differences exist, is BETA normally distributed, following a bell shape, with relatively few strong believers and disbelievers, and many more people close to the scale midpoint? Alternatively, is BETA equally distributed, with about the same number of people sternly believing, sternly disbelieving, and holding moderate beliefs concerning true altruism's existence? Lastly, a bimodal distribution is also possible, with relatively many stern believers and relatively many stern disbelievers, coupled with only a relatively small number of moderates. Figure 3.1 displays the BETA distribution. The figure shows that BETA has a skewed unimodal distribution, with most participants siding somewhat in favor of the existence of true altruism, and fewer participants siding for absolute certainty either way.

Relations with demographics. We wondered whether this prevalence of belief in true altruism (vs. universal egoism) varies as a function of participants' culture, age, or sex. We found no overall differences between German and US participants. However, US participants scored significantly higher than German participants on the positively scored items (albeit the effect size was very small; Cohen, 1977), but the two cultural groups did not differ significantly on the negatively scored items. Further, we found neither age differences nor sex differences on the whole scale or on separate analyses for positively scored and negatively scored items.

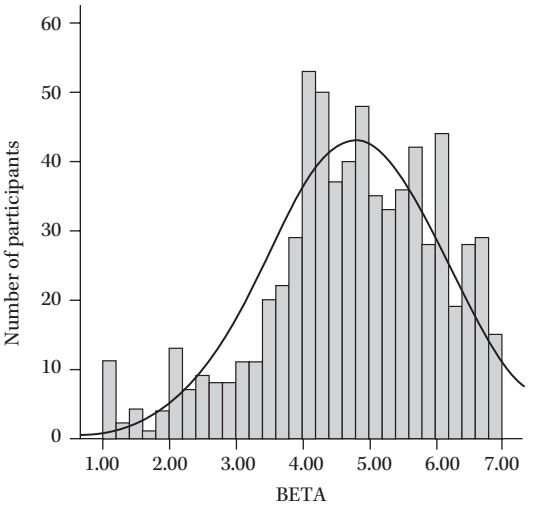

FIGURE 3.1 Distribution of BETA ($N = 729$).

CONCLUSIONS REGARDING RESEARCH QUESTION I

We obtained strong evidence that lay people are more inclined to believe in the existence of true altruism than in universal egoism. This proclivity was evident across participants from two countries (Germans and Americans), almost the entire adult age range, and both sexes. Individual differences on the BETA scale emerged, and they showed a skewed unimodal distribution.

Review on Research Question II: BETA and Other Lay Beliefs

How is BETA related to other lay beliefs? We addressed this question by capitalizing on relevant measures that we included in the above studies. These measures assessed belief in a just world (Dalbert et al., 1987), belief in free will (Paulhus & Carey, 2011), belief in God's existence (Norenzayan & Hansen, 2004), and belief in moral goodness (Rosenberg, 1956; Schuessler, 1982).

We hypothesized that BETA is distinct from the other four lay beliefs, but we also tested or explored additional hypotheses concerning relations between BETA and other lay beliefs. Given that true altruism and moral goodness have been classically considered instantiations of desirable human nature (Aristotle, 1911/350 BC; see also Webster & Saucier, 2013), we expected a positive relation between belief in moral goodness and BETA. Also, given religious references to the existence of true altruism (e.g., 1 Corinthians 13:4–7), we tested whether belief in (a Christian) God is positively associated with BETA. In addition, a positive outlook on the world is associated with belief in a just world (Dalbert, 2009) and with belief in true altruism's existence (Webster & Saucier, 2013).

A positive relation, then, between belief in a just world and true altruism is likely. Finally, philosophers have often described free will and true altruism as desirable human capacities, suggesting a positive association.

We inspected the correlations between the five lay beliefs in the pooled sample from Studies 1–4 ($N = 729$). As anticipated, BETA was distinct from the other four lay beliefs. To begin, we found small, positive relations between BETA and beliefs in a just world, free will, and God's existence (Cohen, 1988). Furthermore, we obtained moderate positive correlations between BETA and the two scales used to assess belief in moral goodness (Cohen, 1988). The relationship between BETA and the other four lay beliefs did not differ between the German and American samples.

CONCLUSIONS REGARDING RESEARCH QUESTION II

Across two cultures, the relations between BETA and other relevant lay beliefs (beliefs in a just world, free will, God's existence, and moral goodness) were small to moderate. Hence, BETA is a unique lay belief. The literature has linked all these other lay beliefs to higher prosociality and moral character. How is BETA, then, linked to moral character, especially in comparison to the other lay beliefs?

Review on Research Question III: BETA and Moral Character

Is belief in the existence of true altruism associated with higher prosociality and more moral character? Gebauer et al. (2015) examined three accounts for such a "BETA-character link." Before reviewing those three accounts, however, we comment on our measurement approach of moral character.

MEASUREMENT OF MORAL CHARACTER

Measuring individual differences in moral character is challenging (see Bleidorn, chap. 5, this volume; Miller, 2013). This broad construct cannot be captured adequately by a single behavior or a handful of behaviors. Hence, we operationalize moral character broadly, drawing on an array of established measures from psychology, sociology, and behavioral economics. In Studies 1–2, we used a measure of retrospectively reported helping (Rushton, Chrisjohn, & Fekken, 1981). In Studies 3–4, we used sociological questions about civic engagement (e.g., belonging to a charity organization, doing volunteer work; Costa & Kahn, 2003; Livingstone & Markham, 2008; Shah, 1998) in addition to the measure used in Studies 1–2. In Studies 5–6 (online studies in the United States and Germany; $N = 682$), we used moral quandaries from philosophy (Bruder & Tanyi, 2014) in addition to the measures from our

prior studies. Finally, in Study 7 (a laboratory study in Germany; $N = 687$), we used three economic-decision games (dictator game [Kahneman, Knetsch, & Thaler, 1986]; ultimatum game [Güth, Schmittberger, & Schwarze, 1982]; trust game [Berg, Dickhaut, & McCabe, 1995]), asked participants to donate some of their participation money to charities (Verplanken & Holland, 2002), and assessed their willingness to volunteer for an Internet study in order to help a graduate student (Coke et al., 1978), in addition to the measures from our prior studies. Thus, our studies included up to eight diverse indicators of moral character. EFAs suggested that, in each of the seven studies, moral character can be adequately assessed by averaging the standardized responses to the measures.

THREE ACCOUNTS FOR THE BETA-CHARACTER LINK

BETA was associated with higher moral character in all seven studies. The size of this BETA-character link ranged between small and medium. But what explains this highly consistent relation between the belief in true altruism's existence and more moral character?

The "spurious-effect account." The BETA-character link may be caused spuriously by other variables with which it is correlated. First, all other aforementioned lay beliefs are related to moral character (as our literature review has shown) and to BETA (as our research has indicated). Therefore, the BETA-character link may be spuriously caused by these other lay beliefs. Second, people differ in the frequency and the degree to which they develop warm, empathic feelings toward others (Davis, 1983). People with a proclivity for empathy should help others more frequently (Eisenberg & Miller, 1987). They should do so because they are more frequently in a state of empathy, which elicits altruistic motivation (Batson, 1991, 2011). Empathic people should also believe more in the existence of true altruism, because their heightened altruistic helping reinforces their belief that true altruism exists (possibly via self-perception processes; Bem, 1967). Hence, the BETA-character link may be spuriously caused by trait empathy. Third, there are potential third variables pertaining to egoistic motivations to help. Examples are individual differences in seeing others as similar to self (i.e., self-other merging; Cialdini et al., 1997; Maner et al., 2002) or in gaining pleasure from helping (Gebauer, Riketta, Broemer, & Maio, 2008). To elaborate, it is, for instance, possible that the belief in true altruism's existence leads people to derive pronounced feelings of pleasure from helping ("look at me, what an altruistic helper I am"), which in turn may lead to exacerbated helping behavior. Finally, communion is a broad personality trait that is conceptually related to prosociality (Gebauer, Sedikides, Lüdtke, & Neberich, 2014). Thus, communion may spuriously inflate the relation between BETA and moral character.

We tested whether the BETA-character link weakened when we controlled for beliefs in a just world, free will, God's existence, and moral goodness.

Pooling the data from Studies 1–4 ($N = 729$) showed that controlling for the other four lay beliefs did not substantially reduce the BETA-character link. In fact, in a model that included BETA and the four other lay beliefs as simultaneous predictors of moral character, only BETA had a significant relation to moral character. If anything, these results suggest that the relation between other lay beliefs and moral character is spuriously caused by the BETA-character link.

We subsequently tested whether the BETA-character link weakened when we controlled for empathic concern. Only Studies 5–7 ($N = 1{,}369$) included empathic concern. Hence, we pooled the data from these three studies. Research has established that empathic concern is one of the most robust correlates of prosocial behavior, and our results replicated this pattern. Specifically, in line with a meta-analysis (Eisenberg & Miller, 1987), we obtained a moderately positive relation between empathic concern and moral character. When simultaneously including BETA and empathic concern as predictors of moral character, the effect sizes of both predictors were somewhat reduced, but both predictors remained significant. Hence, the possibility that the BETA-character link is partly spurious to empathic concern deserves further attention. To this end, we proceeded with examining the longitudinal interrelations between BETA, empathic concern, and moral character. About 75% of Study 7's participants completed the same measures twice with a time interval of about half a year.[1] A full cross-lagged model provided evidence against the idea that the BETA-character link is spurious due to empathy. Specifically, neither did we find a significant longitudinal effect of empathic concern on BETA; nor did we find a significant longitudinal effect of empathic concern on moral character.[2]

Moreover, we tested whether the BETA-character link weakened when we controlled for general self-other merging (Aron, Aron, & Smollan, 1992) and prosocial motivation pleasure (Gebauer et al., 2008). Again, we used the pooled data from Studies 5–7 ($N = 1{,}369$). Simultaneous inclusion of BETA, self-other merging, and prosocial motivation pleasure as predictors of moral character revealed that neither self-other merging nor prosocial motivation pleasure accounted for the BETA-character link. Study 7's longitudinal analyses corroborated these results. Neither self-other merging nor prosocial motivation pleasure had a longitudinal effect on moral character or on BETA.

---

[1] We are currently in the process of collecting a third wave of data for Study 7. Hence, the results of Study 7 should be regarded as preliminary.

[2] In our longitudinal analyses, we operationalized moral character with the average of six (out of our eight) measures. Specifically, two measures showed insufficient temporal stability (the trust game and volunteering in an online study to help a graduate student). This led us to strongly question the suitability of these measures to assess moral character (a character *trait*). At the same time, the stabilities of the other six measures were all comparatively high.

Finally, we tested whether the BETA-character link weakened when we controlled for communion. We had included a measure of communion (Gebauer, Paulhus, & Neberich, 2013) in studies 2, 3, 4, and 7, and we pooled the relevant data ($N = 1{,}267$). Simultaneous inclusion of BETA and communion as predictors of moral character revealed that communion did not account for the BETA-character link. Study 7's longitudinal analyses corroborated these results. Communion had no longitudinal effect on moral character or on BETA.

The "BETA-effect account." In Roman Polanski's movie *Carnage*, Penelope Longstreet ponders a question that is right at the heart of the BETA-effect account. Specifically, she asks, "We are only superficially fair-minded, so why should we be fair-minded at all?" Along the same lines, reflections of BETA disbelievers may go like this: "If all our prosociality is nothing but an indirect (and possibly self-deceptive and other-deceptive) way to benefit ourselves personally, why should I not be outwardly and directly selfish, gaining not only personal benefits, but also authentically expressing my inner self" (cf. Lenton, Bruder, Slabu, & Sedikides, 2013). Consequently, the BETA-effect account predicts a positive relation between BETA and moral character because low BETA lowers moral character.

There is a complementary reason for why BETA may be linked to more moral character. Specifically, people who believe that truly altruistic behavior is indeed part of the human behavioral repertoire may seek to express their altruistic self in an effort to act authentically. As a result of that effort BETA believers may display more moral character than BETA disbelievers.[3]

We examined the BETA-effect account using Study 7's longitudinal data. A cross-lagged panel analysis revealed that BETA predicted moral character six months later, despite our controlling for the temporal stability of BETA and moral character over this time period. The results remained intact even when we controlled for empathic concern, self-other merging, prosocial motivation pleasure, and communion at both time-points. Thus, within the natural limits regarding causal conclusions based on nonexperimental designs, these longitudinal analyses are in line with the BETA-effect account.

The "character-effect account." People high in moral character may self-enhancingly convince themselves of true altruism's existence (cf. Leary, 2007; Sedikides & Gregg, 2008). By doing so, they can feel good about their moral character since it is driven by altruism, which they value (Aristotle, 1911/350 BC; Gebauer, Göritz, Hofmann, & Sedikides, 2012; Gebauer, Joiner, Baumeister, Göritz, & Teismann, 2015). Consequently, the character-effect account predicts a positive relation between

[3] We thank Will Fleeson for suggesting this complementary theoretical explanation for why BETA may lead to more moral character.

BETA and moral character, because high moral character increases belief in true altruism's existence.

We also examined the character-effect account using Study 7's longitudinal data. Our cross-lagged panel analysis revealed that moral character predicted BETA six months later, despite our controlling for the temporal stability of BETA and moral character over this period. The results remained significant even when we controlled for empathic concern, self-other merging, prosocial motivation pleasure, and communion at both time-points.

CONCLUSIONS REGARDING RESEARCH QUESTION III

BETA is positively related to moral character. We obtained this BETA-character link in each of our seven studies. Results indicated that none of a range of third variables could explain the BETA-character link. In contrast, we obtained a longitudinal effect of BETA on moral character, and this effect proved robust even after controlling for empathy, self-other merging, prosocial motivation pleasure, and communion. We also found a longitudinal effect of moral character on BETA, and this effect, too, proved robust when controlling for empathy, self-other merging, prosocial motivation pleasure, and communion. Together, the BETA-character link is not likely to be spurious.

Conclusion

Lay beliefs have received relatively little attention in scientific psychology. Although they were not neglected altogether (Dweck & Leggett, 1988; Lerner, 1965), they received less attention than attitudes, values, goals, or traits. Yet, psychologists have increasingly become aware of the relevance of lay beliefs for human thought, feeling, and behavior. As a result, psychologists have amassed knowledge on key lay beliefs about human nature, such as belief in free will (Baumeister et al., 2009; Carey & Paulhus, 2013), moral goodness (Webster & Saucier, 2013; Wrightman, 1991), and the existence of God (Gebauer & Maio, 2012; Shariff & Norenzayan, 2007). However, one type of belief has received scarce empirical research—namely, belief in the existence of true altruism (BETA). Although true altruism has been discussed interminably for hundreds of years, no research has directly tested lay beliefs in true altruism's existence. In this chapter, we provided an overview of our research on this topic.

We addressed four key questions on BETA. *First*, we asked whether most lay people believe in the existence of true altruism or in universal egoism. We found that, on average, lay people believe in the existence of true altruism rather universal egoism. This pattern held across two cultures (Germany,

the United States), the adult age range, and both sexes. *Second*, we tested whether BETA can be conceptualized as an individual differences variable. We obtained considerable individual differences, which were relatively normally distributed. *Third*, we examined the relation between BETA and other relevant lay beliefs (beliefs in a just world, free will, God's existence, and moral goodness). BETA was consistently related to all these lay beliefs, but the size of the relations was small to moderate. *Finally*, we investigated the relation between BETA and moral character. In each of the seven studies, we obtained a positive relation between BETA and moral character. This BETA-character link occurred robustly despite sampling from two cultures (Germany, the United States) and using a large array of moral character measures, including retrospective self-reports of helping behavior, civic engagement indices, moral quandaries, three different economic games (dictator, ultimatum, and trust), charity donations, and volunteering in a time-consuming task.

We also began to study the nature of the BETA-character link. We tested for the spuriousness of this link due to other lay beliefs, major altruistic and egoistic drivers of prosociality, and trait communion. These third variables did not account for the BETA-character link. Instead, there was a bidirectional longitudinal relation between BETA and moral character. The longitudinal effect of BETA on moral character suggests that future scholarly insight into the existence of true altruism could have an effect on lay people's moral character. For example, if scientific evidence were to unambiguously support universal egoism, one might find an accompanying decline in lay belief in true altruism's existence and, as a result, a decline in moral character among lay people (cf. Vohs & Schoolar, 2008, for a similar argument regarding the belief in free will).[4]

The bidirectional longitudinal relation between BETA and moral character remained intact even when we controlled for empathy, self-other merging, prosocial motivation pleasure, and communion. In fact, in a common cross-lagged panel design, BETA was the only significant longitudinal predictor of moral character; whereas empathy, self-other merging, prosocial motivation pleasure, and trait communion played a little role at best. These findings attest to the usefulness and pervasiveness of the lay belief in the existence of true altruism.[5]

---

[4] Of course, this does not imply that scientific evidence for universal egoism should be concealed or evaluated more critically than evidence for true altruism.

[5] This paper was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. We thank Will Fleeson and Christian Miller for helpful comments on earlier drafts of this chapter.

References

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Aquinas, T. (1917). *The summa theologica* (Vol. 2, Pt. II). (Fathers of the English Dominican Province, trans.). New York: Benziger Bros. (Original work produced 1270.)

Archer, R. L. (1984). The farmer and the cowman should be friends: An attempt at reconciliation with Batson, Coke, and Psych. *Journal of Personality and Social Psychology*, *46*, 709–711.

Aristotle (1911). *The Nicomachean ethics* (D. P. Chase, trans.). New York, NY: E. P. Dutton. (Original work produced 350 BC.)

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*, 596–612.

Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic? In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 65–122). New York, NY: Academic Press.

Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Batson, C. D. (2011). *Altruism in humans*. New York, NY: Oxford University Press.

Batson, C. D., Coke, J. S., Jasnoski, M. L., & Hanson, M. (1978). Buying kindness: Effect of an extrinsic incentive for helping on perceived altruism. *Personality and Social Psychology Bulletin*, *4*, 86–91.

Batson, C. D., Duncan, B., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*, 290–302.

Batson, C. D., Fultz, J., Schoenrade, P. A., & Paduano, A. (1987). Critical self-reflection and self-perceived altruism: When self-reward fails. *Journal of Personality and Social Psychology*, *53*, 594–602.

Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology, Vol. 5: Personality and social psychology* (pp. 463–484). Hoboken, NJ: Wiley.

Baumeister, R. F. (2002). Religion and psychology: Introduction to the special issue. *Psychological Inquiry*, *13*, 165–167.

Baumeister, R. (2008). Free will in scientific psychology. *Perspectives on Psychological Science*, *3*, 14–19.

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, *35*, 260–268.

Baumeister, R. F., Mele A. R., & Vohs, K. D. (2010). *Free Will and Consciousness: How Might They Work?* New York, NY: Oxford University Press.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*, 183–200.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122–142.

Bierhoff, H. W., Klein, R., & Kramp, P. (1991). Evidence for the altruistic personality from data on accident research. *Journal of Personality*, *59*, 263–280.

Bloom, P. (2012). Religion, morality, evolution. *Annual Review of Psychology*, *63*, 179–199.

Bruder, M., & Tanyi, A. (2014). Overdemanding consequentialism? An experimental approach. *Utilitas*, *26*, 250–275.

Cadenhead, C., & Richman, C. L. (1996). The effects of interpersonal trust and group status on prosocial and aggressive behavior. *Social Behavior and Personality*, *24*, 169–184.

Campbell, D. T. (1975). On the conflicts between biological and social evolution and between psychology and moral tradition. *American Psychologist*, *30*, 1103–1126.

Carey, J. M., & Paulhus, D. L. (2013). Worldview implications of believing in free will and/or determinism: Politics, morality, and punitiveness. *Journal of Personality*, *81*, 130–141.

Chiu, C-Y., Hong, Y-Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology*, *73*, 19–30.

Cialdini, R. B. (1991). Altruism or egoism? That is (still) the question. *Psychological Inquiry*, *2*, 124–126.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy-altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, *73*, 481–494.

Cialdini, R. B., Darby, B. L., & Vincent, J. E. (1973). Transgression and altruism: A case for hedonism. *Journal of Personality and Social Psychology*, 9, 502–516.

Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fulz, J., & Beaman, A. L. (1987). Empathy-based helping: Is it selflessly or selfishly motivated? *Journal of Personality and Social Psychology*, *34*, 749–758.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coke, J. S., Batson, C. D., & McDavis, K. (1978). Empathic mediation of helping: A two-stage model. *Journal of Personality and Social Psychology*, *36*, 752–766.

Costa, D. L., & Kahn, M. E. (2003). Civic engagement and community heterogeneity: An economist's perspective. *Perspectives on Politics*, *1*, 103–111.

Dalbert, C. (2009). Belief in a just world. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of Individual Differences in Social Behavior* (pp. 288–297). New York, NY: Guilford Publications.

Dalbert, C., Montada, L., & Schmitt, M. (1987). Glaube an eine gerechte Welt als Motiv: Validierungskorrelate zweier Skalen [The belief in a just world as a motive: Validity correlates of two scales]. *Psychologische Beiträge*, *29*, *596–615*.

Darley, J., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*, 100–108.

Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126.

Dovidio, J. F. (1984). Helping behavior and altruism: An empirical and conceptual overview. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 361–427). New York, NY: Academic Press.

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. A. (2006). *The social psychology of prosocial behavior.* Mahwah, NJ: Lawrence Erlbaum Associates.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273.

Eisenberg, N., & Miller, P. A. (1987). Empathy and prosocial behavior. *Psychological Bulletin*, *101*, 91–119.

Friese, M., & Wänke, M. (2014). Personal prayer buffers self-control depletion. *Journal of Experimental Social Psychology*, *51*, 56–59.

Furnham, A. (2003). Belief in a just world: Research progress over the past decade. *Personality and Individual Differences*, *34*, 795–817.

Galen, L. W. (2012). Does religious belief promote prosociality? A critical examination. *Psychological Bulletin*, *138*, 876–906.

Gallup, G., Jr., & Castelli, J. (1989). *The people's religion: American faith in the Nineties*. New York, NY: Macmillan.

Gebauer, J. E., Göritz, A. S., Hofmann, W., & Sedikides, C. (2012). Self-love or other-love? Explicit other-preference but implicit self-preference. *PLoS One*, *7*, e41789.

Gebauer, J. E., Joiner, T. E., Baumeister, R. F., Göritz, A. S., & Teismann, T. (2015). *Altruistic suicide or escaping a burdened self: Why does suicide ideation increase when one feels like a burden on others?* Manuscript under review, Universität Mannheim, Germany.

Gebauer, J. E., Leary, M. R., & Neberich, W. (2012). Big Two personality and Big Three mate preferences: Similarity attracts, but country-level mate preferences crucially matter. *Personality and Social Psychology Bulletin*, *38*, 1579–1593.

Gebauer, J. E., & Maio, G. R. (2012). The need to belong can motivate belief in God. *Journal of Personality*, *80*, 465–501.

Gebauer, J. E., Paulhus, D. L., & Neberich, W. (2013). Big Two personality and religiosity across cultures: Communals as religious conformists and agentics as religious contrarians. *Social Psychological and Personality Science*, *4*, 21–30.

Gebauer, J. E., Riketta, M., Broemer, P., & Maio, G. R. (2008). Pleasure and pressure based prosocial motivation: Divergent relationships to subjective well-being. *Journal of Research in Personality*, *42*, 399–420.

Gebauer, J. E., Sedikides, C., Leary, M. R., & Asendorpf, J. B. (2015). *Lay belief in the existence of true altruism*. Manuscript in preparation, Universität Mannheim, Germany.

Gebauer, J. E., Sedikides, C., Lüdtke, O., & Neberich, W. (2014). Agency-communion and interest in prosocial behavior: Social motives for assimilation and contrast explain sociocultural inconsistencies. *Journal of Personality*, *82*, 452–466.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*, 367–388.

Hobbes, T. (1651). *Leviathan: Or the matter, form, and power of a commonwealth, ecclesiastical and civil*. London: A. Crooke.

Hoffman, M. L. (1981). Is altruism part of human nature? *Journal of Personality and Social Psychology*, *40*, 121–137.

Hornstein, H. A., LaKind, E., Frankel, G., & Manne, S. (1975). Effects of knowledge about remote social events on prosocial behavior, social conception, and mood. *Journal of Personality and Social Psychology*, *32*, 1038–1046.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–139). New York, NY: Guilford Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59, 285–300.

Kane, R. (2005). A contemporary introduction to free will. New York, NY: Oxford University Press.

Kay, A. C., Gaucher, D., Napier, J. L., Callan, M., & Laurin, K. (2008). God and the government: Testing a compensatory control mechanism for the support of external systems. *Journal of Personality and Social Psychology, 95*, 18–35.

Krebs, D. L. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, *32*, 1134–1146.

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W.-Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. In M. Zanna (Ed.), *Advances in experimental social psychology* (vol. 34, pp. 331–376). New York, NY: Academic Press.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology*, *58*, 317–344.

Lenton, A. P., Bruder, M., Slabu, L., & Sedikides, C. (2013). How does "being real" feel? The experience of state authenticity. *Journal of Personality*, *81*, 276–289.

Lerner, M. J. (1965). Evaluation of performance as a function of performer's reward and attractiveness. *Journal of Personality and Social Psychology*, *1*, 355–360.

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. New York: Plenum Press.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, *85*, 1030–1051.

Livingstone, S., & Markham, T. (2008) Mediating public participation: On the political significance of everyday media consumption. *British Journal of Sociology*, *59*, 351–371.

Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. (2002). The effects of perspective taking on motivations for helping: Still no evidence for altruism. *Personality and Social Psychology Bulletin*, *28*, 1601–1610.

Marsh, H. W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*, 810–819.

McCullough, M. E., & Willoughby, B. L. B. (2009). Religion, self-control, and self-regulation: Associations, explanations, and implications. *Psychological Bulletin*, *135*, 69–93.

Meindl, J. R., & Lerner, M. J. (1983). The heroic motive: Some experimental demonstrations. *Journal of Experimental Social Psychology*, *19*, 1–20.

Miller, C. (2013). *Moral character: An empirical theory*. Oxford, UK: Oxford University Press.

Norenzayan, A., & Hansen, I. G. (2006). Belief in supernatural agents in the face of death. *Personality and Social Psychology Bulletin*, *32*, 174-187.

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, *322*, 58–62.

Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: Measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment*, *93*, 96–104.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York, NY: Guilford Press.

Piliavin, J. A., & Charng, H. (1990). Altruism: A review of recent theory and research. *Annual Review of Sociology*, *16*, 27–65.

Piliavin, J. A., Dovidio, J. F., Gaertner, S. L., & Clark, R. D., III (1981). *Emergency intervention*. New York, NY: Academic Press.

Piliavin, J. A., & Piliavin, I. M. (1973). *The Good Samaritan: Why does he help?* Unpublished manuscript, University of Wisconsin, Madison, WI.

Rokeach, M. (1973). *The nature of human values*. New York, NY: Free Press.

Rosenberg, M. (1956). Misanthropy and political ideology. *American Sociological Review*, *21*, 690–695.

Rosenberg, M. (1957). Misanthropy and attitudes toward international affairs. *Journal of Conflict Resolution*, *1*, 340–345.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rosenhan, D. L., Salovey, P., Karylowski, J., & Hargis, K. (1981). Emotion and altruism. In J. P. Rushton & R. M. Sorrentino (Eds.), *Altruism and helping behavior: Social, personality, and developmental perspectives* (pp. 233–248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rousseau, J. J. (1950). A discourse on the origin of inequality. In *The social contract and discourses*. (G. D. H. Cole, trans.) New York: E. P. Dutton. (Original work published 1755.)

Rubin, Z., & Peplau, L. A. (1975). Who believes in a just world? *Journal of Social Issues*, *31*, 65–89.

Ruether, R. (2005). *Goddesses and the divine feminine: A Western religious history*. Berkeley, CA: University of California Press.

Rushton, J. P. (1976). Socialization and the altruistic behavior of children. *Psychological Bulletin*, *83*, 898–913.

Rushton, J. P., Chrisjohn, R. D., & Fekken, G. C. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, 6, 34–43.

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in newborns. *Developmental Psychology*, *12*, 175–176.

Saroglou, V. (2012). Is religion not prosocial at all? Comment on Galen (2012). *Psychological Bulletin*, *138*, 907–912.

Saroglou, V. (2013). Religion, spirituality, and altruism. In K. I. Pargament, J. Exline, & J. Jones (Eds.), *APA Handbook of psychology, religion and spirituality* (Vol. 1, pp. 439–457). Washington, DC: American Psychological Association.

Saroglou, V. (2014). *Religion, personality, and social behavior*. New York, NY: Psychology Press.

Saroglou, V., Pichon, I., Trompette, L., Verschueren, M., & Dernelle, R. (2005). Prosocial behavior and religion: New evidence based on projective measures and peer ratings. *Journal for the Scientific Study of Religion*, *44*, 323–348.

Schuessler, K. F. (1982). Measuring social life feelings. London, England: Jossey-Bass.

Schwartz, S. H. (1977). Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 221–279). New York, NY, Academic Press.

Shah, D. V. (1998). Civic engagement, interpersonal trust, and television use: An individual level assessment of social capital. *Political Psychology*, *19*, 469–496.

Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*, 803–809.

Sedikides, C. (2010). Religiosity: Perspectives from social and personality psychology. Special issue of *Personality and Social Psychology Review*, *14*(1).

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, *3*, 102–116.

Smith, A. (1853). *The theory of moral sentiments*. London, England: Henry G. Bohn. (Original work published 1759.)

Smith, K. D., Keating, J. P., & Stotland, E. (1989). Altruism revised: The effect of denying feedback on a victim's status to empathic witness. *Journal of Personality and Social Psychology*, *57*, 641–650.

Sober, E. (1988). What is evolutionary altruism? *Canadian Journal of Philosophy*, *18*(sup1), 75–99.

Sober, E., & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Staub, E. (1974). Helping a distressed person: Social, personality, and stimulus determinants. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 293–341). New York, NY: Academic Press.

Stich, S., Doris, J. M., & Roedder, E. (2010). Altruism. In J. M. Doris and the Moral Psychology Research Group (Eds.), *The handbook of moral psychology*. Oxford, England: Oxford University Press.

Stillman, T. F., Baumeister, R. F., Vohs, K. D., Lambert, N. M., Fincham, F. D., & Brewer, L. E. (2010). Personal philosophy and personnel achievement: Belief in free will predicts better job performance. *Social Psychological and Personality Science*, *1,* 43–50.

Stroessner, S. J., & Green, C.W. (1990). Effects of belief in free will or scientific determinism on attitudes toward punishment and locus of control. *Journal of Social Psychology*, *130*, 789–799.

Thomas, G., & Batson, C. D. (1981). Effect of helping under normative pressure on self-perceived altruism. *Social Psychology Quarterly*, *44*, 127–131.

Thomas, G. C., Batson, C. D., & Coke, J. S. (1981). Do Good Samaritans discourage helpfulness? Self-perceived altruism after exposure to highly helpful others. *Journal of Personality and Social Psychology*, *40*, 194–200.

Thompson, W. C., Cowan, C. L., & Rosenhan, D. L. (1980). Focus of attention mediates the impact of negative affect on altruism. *Journal of Personality and Social Psychology*, *38*, 291–300.

Tomas, J., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling*, *6*, 84–98.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Verplanken, B., & Holland, R. W. (2002). Motivated decision making: Effects of activation and self-centrality of values on choices and behavior. *Journal of Personality and Social Psychology*, *82*, 434–447.

Viney, W., Waldman, D., & Barchilon, J. (1982). Attitudes toward punishment in relation to beliefs in free will and scientific determinism. *Human Relations*, *35*, 939–950.

Vohs, K. D., & Schooler, J. (2008). The value of believing in free will: Encouraging a belief in scientific determinism increases cheating. *Psychological Science*, *19*, 49–54.

Webster, R. J., & Saucier, D. A. (2013). Angels and demons are among us: Assessing individual differences in belief in pure evil and belief in pure good. *Personality and Social Psychology Bulletin*, *39*, 1455–1470.

Wrightsman, L. S. (1991). Interpersonal trust and attitudes toward human nature. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. London, England: Academic Press.

4 }

Understanding the Importance and Perceived Structure of Moral Character

Geoffrey P. Goodwin, Jared Piazza, and Paul Rozin

Introduction

Social psychologists have recently been rediscovering the importance of moral character in person perception and evaluation. Much previous theorizing blended the notion of moral character with related notions, such as warmth and communion, thus obscuring its true importance. Our recent research, as well as that of others, has established that moral character is distinguishable from these notions, and is of primary importance in person perception, particularly impression formation. We argue that more attention should now be directed toward resolving three remaining issues: (1) whether moral character and social warmth are distinct dimensions of person perception, alongside competence; (2) how moral character is inferred in naturalistic social interactions; and (3) how perceived moral character is structured. Current evidence is lacking with respect to these questions, particularly the latter two, and so our goal is to outline the current landscape of the literature and suggest some ways forward.

Although the term "character" has multiple meanings even within the domain of person perception (contrast "he's a real character" with "he has real character"), we use the term here to refer broadly to the moral dimensions of a person's personality (see also DeSteno & Valdesolo, 2011; Doris, 2002; Kupperman, 1991; Miller, 2014). Moral character can therefore refer to a whole range of potentially moral traits, for instance, kindness, honesty, fairness, loyalty, and courage, as long as those traits are seen as reflective of the moral side of a person's personality. Although moral traits differ widely, they all appear to involve an orientation that extends beyond narrow self-interest in order to encompass broader interpersonal and societal interests. Accordingly, the moral person is not self-centered, and at times may even be called to behave

in a way that is directly contrary to self-interest (see Schopenhauer, 2009/1841; Singer, 1979, 1995).[1]

At first blush, moral character would seem extremely important to the formation of impressions of other people. Knowing another person's character provides important information about how that person is likely to treat you, including in situations when you are unable to monitor the person's behavior. Moral character also seems fundamental to a person's identity; it may influence the overall impressions we form of someone partly because it is seen as a part of the person's essence (see e.g., Strohminger & Nichols, 2014). However, despite great interest in person perception processes among social psychologists, evidence pertaining to the importance of moral character in person perception was, until recently, fragmented and difficult to assess. One reason for this, as will be seen presently, is that social psychologists have been reluctant to theorize explicitly in terms of the notion of moral character, preferring instead more nebulous notions such as "warmth" or "communion."

In what follows, we first review and critique existing research that has prioritized the notion of warmth in person perception, before describing our recent research that has afforded moral character a much more prominent role. We conclude by outlining a series of unresolved questions regarding moral character in person perception.

Review and Critique of Existing Social Cognition Research

Part of the reluctance to emphasize moral character may owe to the strong situationism that dominated social psychology in the latter part of the twentieth century. Hartshorne and May's (1928) famous study, which revealed little cross-situational consistency in the expression of children's moral traits, provided an early impetus for skepticism regarding the existence of truly "global" character traits.[2] So, too, did later seminal studies that demonstrated the power of situational forces on expressions of morally relevant behaviors (e.g., Asch, 1952; Darley & Batson,

[1] It may be too narrow to conceive of moral character as solely about being "other-regarding," as Schopenhauer (1995/1841) seemed to conceive of it. Such a conception prioritizes certain moral character traits, such as compassion or kindness, which orient a person to others' immediate welfare (and suffering). Although they are undoubtedly important, these virtues may sometimes conflict with more austere, principle-focused virtues, such as honesty, trustworthiness, or loyalty, and there is no immediately obvious reason why these principle-based virtues should take a back seat. We therefore think that it is more conservative to think of moral character traits as involving the prioritizing of any sort of interpersonal or societal interest that extends beyond self-interest, rather than "other interest" specifically.

[2] "Global" traits are traits that are (1) consistently exhibited across different situations (something Hartshorne and May's research questioned), and (2) reliably exhibited within the same situation, i.e., they are stable within the same situation (see Doris, 2002; Miller, 2010).

1973; Darley & Latané, 1968; Deutsch & Gerard, 1955; Isen & Levin, 1972; Milgram, 1963; Regan, Williams, & Sparling, 1972). These studies seemed to some theorists to undercut the whole notion that personality traits (let alone moral character traits) are consistent across situations, and led some philosophers to conclude that global character traits are nonexistent (e.g., Doris, 2002; Harman; 1999; though for countervailing arguments, see Kamtekar, 2004; Kupperman, 2001; Miller, 2003, 2009, 2010; Sabini & Silver, 2005). Moreover, even before many of the classic situationist studies emerged in the 1960s and 1970s, there was already considerable skepticism within social psychology about the notion of moral character, stemming in part from Gordon Allport's insistence that "character" was too value-laden a term to be studied scientifically (see Nicholson, 1998). Allport (1937) maintained that only the more value-neutral construct of "personality" could be studied objectively.

Allport's theorizing can be seen as exerting a strong influence on the study of personality, arguably leading to the modern conceptualization of personality in mainly non-moral terms (e.g., in terms of the Big Five; McCrae & Costa, 1987). Although there are arguably some morally relevant elements to the Big Five dimensions of agreeableness and conscientiousness, it has only been recently that an explicitly moral dimension, honesty/humility, has been posited as a distinctive additional feature of personality, as part of the HEXACO model (see e.g., Ashton & Lee, 2001; Ashton, Lee, & Goldberg, 2004; Ashton et al., 2004).

Finally, we speculate that the notion of character may have been seen as unworthy of scientific study owing to its being seen as irreducibly subjective and value-based: a vestige of logical positivism in philosophy and of behaviorism in psychology.

While our focus is on person perception rather than personality, an indirect consequence of the general skepticism among psychologists regarding the existence of global character traits seems to have been their reluctance to couch theories of social cognition and person perception in the language of character as well. The person perception literature has instead been dominated by two-dimensional theories which argue that two roughly orthogonal dimensions, "warmth" and "competence," structure the impressions we form of other people and the social groups they belong to (e.g., Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002). These models arguably include some aspects of morality within the warmth dimension (e.g., kindness), but they obscure the full contribution of moral character to person and group perception.

Two-dimensional models of this sort originated with Rosenberg, Nelson, and Vivekananthan's (1968) famous study which showed that the way in which people think about traits can be adequately organized along two dimensions, one representing social goodness-badness, and the other representing intellectual goodness-badness—the dimensions which later came to be referred to

as "warmth" and "competence," respectively. Less well known among social psychologists, however, is that the evidence from some of Rosenberg's later investigations (e.g., Rosenberg & Olshan, 1970) was more consistent with a three-dimensional model, in which these two dimensions were complemented by a third more generally evaluative dimension (see Landy, Piazza, & Goodwin, 2015).

Social psychologists have long been enamored with the notion of warmth, going at least as far back as Asch's seminal studies on person perception. Asch (1946) showed that adding the single terms "warm" or "cold" to a list of traits describing a target person radically transformed the overall impressions people formed of the target's other traits. The addition of "warm-cold" produced highly more polarized impressions of those other traits, and did so to a greater degree than did other valenced trait pairs, such as "polite-blunt." Asch was particularly impressed by the extent to which the warmth-coldness dimension affected impressions of other, seemingly distally related traits. This led him to see the warmth-coldness dimension as a particularly central aspect of person perception, a view that strongly influenced later theorizing (see e.g., Kelley, 1950; Rosenberg et al., 1968; Williams & Bargh, 2008).

However, while undoubtedly potent in some respects, the notion of warmth contains an important ambiguity. What exactly does it mean for somebody to be warm? On the one hand, warmth seems to connote social gregariousness—being extroverted, affable, playful, funny, sociable, and so forth—without necessarily implying much about a person's moral character. But, on the other hand, warmth also seems to connote kindness, compassion, gratitude, forgiveness, and so forth—traits that are undeniably moral. Authors who favor two-dimensional models have tended to operationalize warmth in a variety of different ways, reflecting this bifurcated meaning (as noted by, Bauman & Skitka, 2012; Brambilla, Rusconi, Sacchi, & Scherabini, 2011; Brambilla, Rusconi, Sacchi, Scherabini, & Yzerbyt, 2012; Goodwin, Piazza, & Rozin, 2014; Landy et al., 2015; Leach et al., 2007). Some proponents of two-dimensional theories have argued explicitly that warmth includes morality (see e.g., Cuddy et al., 2008, p. 65). Yet, in these models, warmth is also clearly seen as including sociability as well (Fiske et al., 2002), and sometimes sociability is emphasized as the dominant component of warmth (see e.g., Kervyn, Bersieker, & Fiske, 2012; Lin, Kwan, Cheung, & Fiske, 2005).

There are two problems with this dual-aspect account of the warmth dimension. First, at the conceptual level, it is not clear that many central moral traits actually involve much warmth at all. While it has been posited that being honest, fair, trustworthy, and righteous, for instance, are warmth traits (Cuddy et al., 2008, p. 65), there is little evidence to support this contention, and it seems questionable on its face. It would seem entirely possible for a person to embody each of these traits without the affection, enthusiasm, or cordiality that characterizes what it means to be "warm" (Warm, n.d.). A second, more

empirical problem, is that the conflation of morality and warmth downplays the important role that moral character traits have in person perception and evaluation.

Several lines of recent research have begun to emphasize the importance of moral character information in these areas. For instance, there have been several demonstrations of the special importance of the particular moral trait of trustworthiness. People's beliefs about another person's trustworthiness have been shown to influence their likelihood of cooperating with that person (Burnham, McCabe, & Smith, 2000; van't Wout & Sanfey, 2008). Trustworthiness is also regarded as being extremely important for a wide variety of interpersonal relationships (Cottrell, Neuberg, & Li, 2007), and it appears to be more important than either competence or sociability information in positively affecting impressions of one's in-group (Leach et al., 2007).[3] Trustworthiness is also a core part of Americans' prototype of an ideal US president (Kinder, Peters, Abelson, & Fiske, 1980). Finally, appraisals of a person's trustworthiness can be made extremely rapidly, sometimes within 100 milliseconds of exposure to another's face, attesting to the automaticity (and therefore, arguably, the importance) of these appraisals (Willis & Todorov, 2006).

There has also been an increasing trend among some two-dimensional theorists to refer to the social axis of their models in terms of the notion of "communion," rather than warmth. According to one definition, communion refers to "a person's striving to be part of a community, to establish close relationships with others, and to subordinate individual needs to the common good" (Abele, Uchronski, Suitner, & Wojciszke, 2008, p. 1204). This is arguably a more characterological notion than is warmth, particularly in its emphasis on subordinating individual needs to the common good. Thus, although two-dimensional theorists have sometimes seen communion as almost identical with warmth (see e.g., Abele, Cuddy, Judd, & Yzerbyt, 2008), its increasing prominence suggests a shift among two-dimensional theorists toward a greater focus on moral character, though within the constraints of existing two-dimensional theories.[4]

Finally, some theorists have explicitly conceived of the notion of "sociability" as being separate from that of morality, and have emphasized the dominant role of morality in social cognition. Leach et al. (2007) argued that moral traits are particularly relevant to people's evaluations of their own in-groups, more so than competence or sociability traits. Similarly, Brambilla et al. (2012)

---

[3] Although Leach et al. (2007) conceptualize these findings as speaking to the general importance of "morality" information, it is more accurate to treat them as speaking to the importance of trustworthiness, specifically, since the three traits used to instantiate morality were all closely related to this notion of trustworthiness (e.g., honesty, sincerity, trustworthiness).

[4] It is worth noting, however, that even the notion of communion is not identical with that of moral character. A person who is highly focused on communion might be inclined to act in popularity- or inclusion-seeking ways that could at times conflict with moral imperatives (see also Goodwin et al., 2014).

argued that impressions of social groups tend more strongly to be determined by evaluations of those groups' moral traits than by evaluations of their sociability traits. Meanwhile, Brambilla et al. (2011) showed that people tend preferentially to select information about moral traits when forming an overall impression of another person.

Each of these investigations represents an important precursor to our own work on the role of moral character in person perception. Our focus was primarily on comparing moral character with warmth, as these terms are conceived in everyday terms.[5] Building on the insights of prior theorists, we hypothesized that moral character information is more important than warmth information in determining the impressions people form of others in their daily lives.

While prior investigations alluded to this broad conclusion, none had shown this in a comprehensive way. Partly, this is because the range of moral traits investigated in prior studies has often been quite narrow. In Leach et al.'s (2007) and Brambilla et al.'s (2012) studies, for instance, only three closely synonymous traits were investigated: honesty, sincerity, and trustworthiness. It therefore remained uncertain from these studies whether the documented importance of moral information would generalize to a much wider range of moral traits, including those related to compassion, justice, loyalty, dedication, humility, courage, and so forth. Additionally, prior studies have not chiefly focused on how moral character and warmth information contribute directly to the actual formation of global impressions about other people. Leach et al. (2007) focused on impressions about one's own in-groups; Brambilla et al. (2012) focused on impressions of social groups; and Brambilla et al. (2011) focused on the selection of information for impression formation. Prior studies that have focused on impression formation have compared morality only with competence information. Importantly, these studies have revealed that moral information tends to be more important in determining the impressions we form of others than competence information (Wojciszke, Bazinska, & Jaworski, 1998).[6] But they do not reveal whether moral character is more important than warmth.

---

[5] The everyday conception of "warmth" is likely to reflect the bifurcated meaning described above. But we imposed no a priori constraints on this. Instead, as will be seen presently, we relied on an empirically driven, bottom-up measurement procedure to assess traits' relevance to these two concepts, morality and warmth.

[6] We also included competence in some of our investigations, though we did not focus on it primarily. The notion of competence within two-dimensional theories has generally been thought to reflect a person's underlying skills and abilities (e.g., Cuddy et al., 2008; Fiske et al., 2007) or, relatedly, a person's effectiveness in accomplishing their goals (see e.g., Wojciszke, Bazinska, & Jaworski, 1998). More recent two-dimensional conceptions refer instead to the notion of "agency" (e.g., Abele et al., 2008), which is related to but slightly different from the notion of competence (see Cuddy et al., 2008). Whichever conception is used, however, competence is regarded as less important than warmth (e.g., Abele & Bruckmüller, 2011; Fiske et al., 2007; Wojciszke, Bazinska, & Jaworski, 1998). Competence is thought to augment or amplify warmth, which is primary.

Review of Our Recent Work on the Role of Moral Character in Person Perception

Accordingly, in our own work, we have comprehensively investigated the respective roles of moral character and warmth in person perception and evaluation. Our starting point was to undertake a large-scale, bottom-up investigation of the extent to which various traits are seen as being relevant to moral character and to warmth, respectively (Goodwin et al., 2014). Traits have often been classified as instantiating broad personality dimensions without adequate justification (though see Abele & Wojciszke, 2007; Wojciszke, Dowhyluk, & Jaworski, 1998). For instance, the designation of "morality," "trustworthiness," and "honesty" as warmth traits (e.g., Cuddy et al., 2008, p. 65) is questionable, because each of these traits could seemingly be enacted without warm or affectionate feeling.

We began by recruiting a large set of American adults (recruited online via Amazon's Mechanical Turk platform [https://www.mturk.com/mturk/welcome]) to provide ratings for 170 different personality and ability traits on several dimensions of interest (Goodwin et al., 2014, Study 1). For each trait, we asked for ratings of how useful having information about the trait would be in determining: a person's character (specifically, whether it is admirable), their goodness or badness, their morality or immorality, and their warmth or coldness, among other things. The traits were selected from eight different categories: goodness/badness, strength/weakness of will, ability/lack of ability, and positive/negative (non-moral) personality, so as to represent a wide variety of different aspects of people's personality and abilities. We were particularly attentive to including traits that reflected various aspects of morality, warmth, and competence.

Across all 170 traits, the raw correlation between moral character usefulness (comprising a combined index of the admirable character, goodness, and morality questions) and warmth usefulness was high: .76. However, what this high overall correlation partly reflects is that many traits within the overall set are not especially relevant to either moral character or warmth. Such traits have very similar, low ratings in terms of their usefulness for judging either construct, thus inflating the overall correlation. Illustrating this, when the range of traits is restricted to the 50 traits that were rated most highly on their usefulness for judging moral character (the composite score), the raw correlation between overall moral character usefulness and warmth drops to .13 and is not significant, $p = .38$. But, among the 50 traits that were rated as least useful for judging moral character, the relation between character relevance and warmth relevance is quite substantial, $r\ (48) = .54$, $p = < .001$, reflecting the fact that these traits (which include many ability/competence traits) have little to do with either character or warmth (the mean usefulness for judging moral character among such traits was $M = 3.65$, while the mean usefulness for

judging warmth was $M = 3.97$ (rated on a 1–9 scale). Such traits tend to be rated as *similarly* low on these two dimensions, thus leading to a higher correlation between the two dimensions among these traits. This therefore helps explain the surprisingly high correlation between character and warmth across all 170 traits. What is most important is that warmth is not strongly related to moral character among the most character relevant traits.

In sum, there is some relation between moral character and warmth, but there is also clear divergence. As a consequence, different groups of traits can be organized according to their differential relevance to each construct. Some traits are viewed as highly relevant to both character and warmth, for example, kindness, humility, cooperativeness, gratefulness, and empathy; some traits are viewed as highly relevant only to moral character but not to warmth, for example, honesty, trustworthiness, loyalty, justness, courage; some traits are viewed as highly relevant only to warmth but not to character, for example, sociability, happiness, agreeableness, humor, and playfulness; while still other traits are viewed as relevant to neither construct (primarily ability traits), for example, intelligence, athleticism, creativity, organization, and cleverness (see further, Goodwin et al., 2014). Drawing upon these empirically demarcated clusters of traits, we next aimed to determine which sorts of traits most strongly determine the overall impressions and evaluations that people form of other individuals.

In a first, correlational study (Goodwin et al., 2014, Study 3), we asked people (in this case, undergraduate students) to think of specific individuals from their lives who fit specific roles—a person they admired, a friend they liked, a parent or parental figure, a person they disliked, a person they did not respect, as well as US presidents George W. Bush and Barack Obama. Participants rated these social targets in terms of the extent to which each target possessed traits from each of the categories described above (high morality, high warmth traits; high morality, lower warmth traits; high warmth, lower morality traits; low morality, low warmth traits), as well as to indicate the positivity or negativity of their overall impressions of each target. That is, participants were asked to indicate "how positive or negative is your overall impression of this person?" on a 9-point scale ranging from *very negative* to *very positive*. Regression analyses for each target indicated that, overall, the moral traits were a far more potent predictor of such global impressions than were the warmth traits. Somewhat to our surprise, the traits that were rated high on morality and lower on warmth (e.g., honesty, loyalty, justness) were the most important, predicting independent variance in global impressions across all seven targets, and thereby playing a more important role than even the high morality, high warmth traits (e.g., kindness, humility, cooperativeness). By contrast, the high warmth, lower morality traits (e.g., sociability, humor, playfulness) tended not to predict global impressions once ratings on the moral character traits were controlled for. Therefore, once the relation between moral character and

warmth is statistically accounted for, traits that reflect purely moral information continue to predict global impressions very strongly, whereas traits that reflect purely warmth information do not.

Two later experimental studies, conducted with American adults recruited online (Mechanical Turk), revealed a very similar overall picture (Goodwin et al., 2014, Studies 5–6). In these studies, we experimentally manipulated the traits possessed by hypothetical targets that were to fulfill a number of different social roles in the participants' lives. The experimental manipulations, though slightly different across the two studies, independently crossed moral character and warmth. Thus, one target possessed a set of traits indicating that he or she was high on both morality and warmth, a second target was high on morality only, a third target was high on warmth only, and a fourth target was high on neither dimension. Because we were particularly interested in examining possible boundary conditions on the importance of moral character information, we sampled broadly from a variety of social roles, so that participants were asked to imagine the target fulfilling 12 different social roles: boss/ supervisor; family relative (cousin, aunt, or uncle); close friend; social acquaintance (at work, school, or place of worship); surgeon (who will be operating on you); co-worker (who you have been assigned to work closely with); your daughter's fiancé; long-term romantic partner; cashier at grocery checkout counter; parent (father, mother, or parental guardian); your child's primary school teacher; and judge (presiding over legal proceedings you are involved in). For each role, participants indicated how positive or negative they would feel about the described person fulfilling the role in question, and the extent to which they would be pleased about the target person fulfilling the role (both were assessed on 100-point scales, and we aggregated these two measures). Although we expected that moral character traits would generally be more important in determining people's overall impressions, we suspected that warmth traits may be particularly important for some of the roles we had chosen.

Moral character traits were once again far more powerful in driving people's global impressions of the targets than were social warmth traits. In both studies, moral character traits influenced global impressions significantly more strongly than warmth traits for 9 out of the 12 roles (the exceptional roles were slightly different across the two studies). And in neither study was there a single social role for which the warmth traits were significantly more potent than the moral character traits. Even for roles where warmth might be expected to come to the fore in driving impressions (e.g., a social acquaintance), moral character information was still seen as being just as important as warmth information. Furthermore, in both studies, moral character drove global impressions more strongly as the roles themselves were rated as more important. These results do not simply reflect peoples' theories that moral traits *should* matter to their evaluations of others: the key trait manipulations in these studies were between-subject ones (with the role manipulation being

within subjects). This design feature thus conceals the true purpose of the study from the participants, and obviates a concern that subjects may have been providing socially desirable responses that would not reflect their true judgments.

Thus, in sum, these experimental studies show that moral character information predominates relative to warmth in determining the overall impressions we form of other individuals. And because these studies are experimental, they reveal direct causal evidence for the influence of character on impressions, which cannot be obtained from correlational studies. While there are some boundary conditions on the dominant influence of character, these limiting conditions seem quite circumscribed.

To corroborate these findings in a different, and more naturalistic setting, we examined how moral character information may figure in summary accounts of people's lives, namely, in their obituaries (Goodwin et al., 2014, Study 7). We sampled over 200 obituary notices from the "notable deaths" archive in the *New York Times*. Supporting the idea that character rises to the fore in summary accounts of people's lives, the obituaries contained substantially more information about moral character than social warmth.

Moreover, the character information strongly determined the impressions people formed of the deceased individuals. To show this, we asked American adults (from Mechanical Turk) to each read three obituaries (different sets of three for each participant sampled from the overall set of more than 200) and to provide an overall valenced impression of each person. The information conveyed about moral character by each obituary was a stronger predictor of overall impressions than the information conveyed about social warmth (though both sorts of information independently predicted impressions in a regression). This result held even when statistically controlling for the fact that more information was conveyed in the obituaries about moral character. Once again, therefore, moral character trumped warmth in predicting global impressions. The clear conclusion from the correlational, experimental, and archival studies we have conducted on this topic is that moral character predominates relative to warmth in contributing to valenced global impressions we form of other people.

Although it was not our main goal to compare moral character information with competence information, past research has clearly established that moral character information is more important than competence information (Wojciszke, Bazinska, & Jaworski, 1998). This therefore leads us to think that moral character is indeed the predominant determinant of global impressions.

Moral Character Is Evaluated Based on a Person's Weakest Link

The way in which moral character traits combine to produce global impressions appears not to be strictly additive, however. Instead, moral character

traits are negatively dominant (see Baumeister, Bratslavsky, & Finkenauer, 2001; Rozin & Royzman, 2001), so that the presence of a negative or immoral trait has a stronger impact on our impression of someone than does the presence of a positive moral trait. Past evidence has shown that people are more inclined to make "correspondent" inferences—that is, inferring that a single piece of a person's behavior reflects a corresponding disposition—from a person's immoral behavior than from their moral behavior. In one study, people discounted the role of situational inducements or pressures when a person acted immorally much more so than when a person acted morally (Reeder & Spores, 1983). That is, someone who acted immorally (e.g., by stealing money) because of a powerful situational inducement was still seen as deeply immoral, whereas a person who acted morally (e.g., donating money) because of a powerful inducement was not necessarily seen as especially moral. Relatedly, initial impressions of a person's immorality tend to be more entrenched than initial impressions of their morality. A large number of compensatory moral behaviors are needed to overcome an initial impression of another person's immorality, whereas sometimes a single immoral behavior is enough to overturn an initial impression of a person's high morality (Reeder & Coovert, 1986).

Given these prior findings, we hypothesized that this negativity dominance should also apply at the level of individual traits. That is, an individual who has a mixture of positive and negative moral traits (e.g., a person who is kind but dishonest, or unkind but honest) should be judged more negatively than an individual who has a mixture of positive and negative non-moral personality traits (e.g., a person who is optimistic but unsociable, or pessimistic but sociable), and also more negatively than an individual who has a mixture of positive and negative ability traits (e.g., a person who is creative but illogical, or a person who is uncreative but logical; see Skowronski & Carlston, 1987, for a related prediction). We have recently corroborated this prediction in a series of studies (Piazza, Goodwin, & Rozin, 2015). Incongruent pairs of moral traits (e.g., kind-dishonest) led to negative overall impressions, whereas incongruent pairs of non-moral personality traits (e.g., optimistic-unsociable) or ability traits (creative-illogical) led either to positive or neutral overall impressions. By contrast, congruent pairs of moral traits (e.g., a person who is both kind and honest) led to somewhat more positive impressions than did congruent personality (e.g., optimistic-sociable) or ability (creative-logical) pairs. And although moral traits tend to have more extreme valences, both positive and negative, than non-moral traits, the negativity dominance effect that we observed cannot be attributed to this alone. When the extremity of the trait pairs' valences were statistically controlled, or experimentally equated for, the negativity dominance of incongruent moral traits still remained (though the relative difference in the impressions caused by incongruent moral traits and incongruent non-moral traits was markedly reduced). In sum, what this suggests is that, particularly in the domain of moral character, a person is judged by their "weakest link."

This weakest link result has some important practical implications. Although proponents of positive psychology have sometimes exhorted people to focus on, and aim to accentuate, their "signature strengths" (Seligman, 2002), the weakest link phenomenon suggests that this strategy may backfire in the moral domain. It may instead be more important, and more prudent, to focus on areas of moral deficit (see also Bermant, Talwar, & Rozin, 2011).

Moral Character Is Seen as Fundamental to Identity

Beyond the role that character plays in determining global impressions, moral character information is also seen as particularly fundamental to a person's identity. We asked people to indicate, for each of 80 positively valenced traits, the extent to which the trait is fundamental to a person's identity, as well as a range of other features, including how changeable and controllable the trait is, and how much it falls under a person's own responsibility (Goodwin et al., 2014, Study 2). Traits that were highly relevant to moral character (as determined by our set of prior ratings) were seen as much more fundamental to identity than traits that were highly relevant to warmth.

Other researchers have also recently converged on a similar conclusion about the central role of moral character information in defining a person's identity. For instance, Strohminger and Nichols (2014) asked participants to consider hypothetical brain transplant cases. In one case, the brain transplant caused a person to lose his ability to recognize objects (agnosia); in another, it caused him to lose all of his desires; in another, it caused him to lose all his autobiographical memories, and in one further case, it caused him to lose his moral conscience. The greatest change to identity was attributed when the man lost his moral conscience. These results replicated when participants considered a different way that a person's brain had changed (via a magic pill). Thus, using these quite different methods, these researchers arrived at the same conclusion: moral character is seen as most fundamental to a person's identity.

Based on these findings, one might expect that moral character traits would be seen as especially indelible and resistant to change. After all, if a trait is fundamental to identity, it would seem strange for it also to be highly variable over time. Arthur Schopenhauer and other philosophers have held this view, regarding moral character as essentially unchangeable, with a person's moral or immoral behavior simply revealing their pre-existing character (e.g., Schopenhauer, 1995/1941). However, contrary to this idea, our participants regarded moral character traits as among the most changeable, and controllable. Such traits were also seen within the domain of a person's own responsibility. Moral character was therefore seen as both fundamental and highly changeable (Goodwin et al., 2014).

The relation between moral character traits' fundamentalness to identity and their importance in determining global impressions is currently unknown. It

could be that character traits powerfully shape global impressions because they are seen as fundamental to identity. To the extent that a particular trait is seen as fundamental to identity, people are likely to pay more attention to it when forming an overall impression of someone. But it could also be that causality operates in the reverse direction as well, so that character traits' fundamentalness to identity stems from their role in shaping global impressions, which in turn may be based on their importance in signaling a person's suitability for cooperation and affiliation (Strohminger & Nichols, 2014). More research is needed to address this issue.

Open Questions

Although we have learned much about the role of character in person perception in recent years, several key questions remain unresolved. Here, we focus on three questions that we see as being especially pertinent: (1) Is morality properly conceived of as a third dimension of person and group perception, separate from both competence and warmth? (2) How is morality and warmth information actually obtained in ordinary social interactions? (3) What is the perceived structure of moral character?

IS MORALITY PROPERLY CONCEIVED OF AS A THIRD DIMENSION OF PERSON AND GROUP PERCEPTION?

As we have argued, existing two-dimensional theories of person and group perception can fruitfully be refined by separating traits that are relevant to moral character from traits that are relevant to social warmth (or "sociability"; e.g., Brambilla et al., 2011; Brambilla et al., 2012; Brambilla, Sacchi, Pagliaro, & Ellemers, 2013; Leach et al., 2007). It is now clear that moral traits are far more important with respect to impression formation than are warmth or sociability traits. But what remains unresolved is whether the warmth dimension of existing two-dimensional models is best replaced by a moral dimension, preserving a two-dimensional framework, or whether, instead, morality and warmth (or sociability)[7] are best conceived as separate but correlated dimensions of person and group perception, alongside competence.

There may be no single definitive answer to this question because the number of dimensions that underlie these social-cognitive processes may shift depending on the context, on the social targets in question, and on the way in which those targets are evaluated (e.g., the traits they are assessed on). It may be unrealistic to think that only one framework is ideal for all circumstances. Nonetheless, the prevailing assumption of existing two-dimensional theories

[7] Hereafter, we will refer to this dimension as "sociability," in order to remain most consistent with existing literature.

is that there is at least some kind of default regularity underlying person and group perception, which can be captured in a two-dimensional model. We think this notion of a default social-cognitive framework is quite reasonable. However, what we wish to question is whether this default regularity is best conceived in terms of two dimensions, as opposed to three, or more.

Surprisingly, there have been few direct tests of the adequacy of existing two-dimensional frameworks within social cognition. Almost all of the many empirical investigations that have relied on a two-dimensional framework have simply assumed this framework as a starting point for exploring a particular domain of interest, without directly testing its adequacy. This is not to criticize such studies: their focus has not been to test two-dimensional theories directly. But it is to point out that little evidence exists that is pertinent to this question, and that the large mass of studies that assume a two-dimensional framework cannot be taken as evidence in favor of that framework, as opposed to some other, possibly three-dimensional, framework. Rosenberg et al.'s (1968) foundational studies are of a genuinely exploratory nature, and do seem to provide evidence that is consistent with a two-dimensional framework, though their evidence is also consistent with a three-dimensional framework. And, as we pointed out earlier, some of Rosenberg's other investigations strongly suggest something more like a three-dimensional framework (e.g., Rosenberg & Olshan, 1970).

Accordingly, we have begun to investigate this issue in a systematic way. In two recent studies, we asked participants to rate a number of target individuals (e.g., "someone you respect," "someone you dislike"; Study 1), as well as a number of target groups (e.g., "a group the average American fears"; Study 2) in terms of a set of six moral character traits (e.g., moral, principled, honest, trustworthy, fair, responsible), six warmth (or sociability) traits (e.g., sociable, warm, friendly, easy-going, extroverted, playful), and six competence traits (e.g., competent, capable, intelligent, effective, skillful, talented; Landy et al., 2015). Our question was how ratings of these traits would be organized by a factor analysis. This simple method enables a critical test of competing frameworks. If a two-dimensional framework is correct, we would expect two distinct factors to emerge (i.e., morality and warmth/sociability traits loading together, separate from competence traits); whereas, if a three-dimensional framework is correct, three factors should emerge. In fact, a three-dimensional factor solution emerged for every single target that participants rated—six individual targets (Study 1) and nine group targets (Study 2). The factor structure was clearly identifiable as reflecting the three dimensions of morality, warmth/sociability, and competence. Moreover, participants' mean ratings of the various individuals and groups on the morality and warmth factors were significantly different from each other the vast majority of the time, providing further corroboration of the distinctness of these two dimensions. The three dimensions that emerged were not totally orthogonal, and were all correlated with each other to

a moderate extent, which is not surprising in light of the fact that each dimension is positively valenced. However, of critical importance to our current theorizing, these correlations were not larger between the morality and warmth/sociability factors than they were between the morality and competence factors. Morality is no more closely related to warmth than it is to competence, thus showing that these two dimensions are not redundant with one another.

This evidence therefore points toward the superior descriptive adequacy of a three-dimensional as opposed to a two-dimensional model, at least for these data. Owing to the caveats just mentioned, however, we are cautious in interpreting these data as providing definitive general evidence for a three-dimensional model. Which dimensional structure emerges may depend in part on the targets evaluated and the traits the targets are evaluated on. Nonetheless, it should be noted that we included more traits than is typical within the two-dimensional literature, and we can be quite confident that a framework including *at least* three dimensions best fits these data.

Going forward, we believe that further critical tests of the adequacy of two- and three-dimensional models (or still-more-complex models) are needed to fully explore a variety of social-cognitive processes, including person perception and evaluation, group perception, stereotyping, and so forth. Doing so will allow more robust conclusions to be drawn about the true dimensional structure underlying these facets of social cognition.

HOW IS MORALITY AND WARMTH INFORMATION ACTUALLY OBTAINED IN ORDINARY SOCIAL INTERACTIONS?

How do people gauge whether another person has good moral character? How do they tell whether someone is warm or sociable? In the studies we have conducted, information about morality and warmth was conveyed in the form of trait information. But, in real life, information about others' traits often does not come neatly packaged in trait lists. In such cases, it must be inferred from actual behavior (see, e.g., Malle & Holbrook, 2012). How do people make these inferences about morality and warmth?

Gauging another person's warmth should arguably be easier than gauging their morality, since warmth is a much more visible trait dimension and is therefore almost immediately apparent in the course of a social interaction. In a sense, warmth can be socially "performed." In order to convince you that I am a warm person, I need to act warmly toward you, perhaps over a series of interactions.

In contrast, a person's moral character is often much less immediately observable, and should take substantially more time to establish with any reasonable certainty. One reason for this is that opportunities to display moral character traits do not present themselves as often in daily life as do opportunities to display warmth traits. Take honesty, for instance. Of course, it is

possible to lie in any conversation, but cases in which lying would truly be beneficial for the person lying do not occur especially frequently—the right opportunity must present itself (see also Skowronski & Carlston, 1987). Similarly, a person's loyalty to a friend or confidant is not usually tested in the course of a typical day. It takes some special circumstance to provide the adequate test. The same general point is true for many other moral character traits (e.g., fairness, courage).

Admittedly, there are several important moral traits that can be expressed more frequently. For instance, a person could theoretically display charitableness, kindness, or compassion almost without limit—by ceaselessly working for the benefit of others, donating money to charity, and so on (Singer, 2009). This applies to many character traits that are relevant for the fulfillment of what Kant described as 'imperfect' duties—that is, those duties that admit of exceptions and which require actions that are strongly recommended but that do not have to be performed all of the time. Such duties typically involve selfless acts of charity or compassion (see Trafimow & Trafimow, 1999), and allow for some choice about how to fulfill them (e.g., which charities to donate to, and how much of the time). However, unrestrained compassion or charity is rare, and it is likely that most ordinary inferences of another person's compassion (or its absence) do not come from observing ceaseless compassion. Rather, such inferences arise from situations in which discrete opportunities to display compassion arise, or from situations that pit competing values against each other. Thus, even for many traits that enable the enactment of imperfect duties, opportunities to display such traits in the ordinary course of daily life may also depend on circumstance.[8] A second complication in inferring moral character traits is that highly public displays of moral behavior are inherently ambiguous as to their underlying motivation, and therefore not always diagnostic. Because there is a stiff social penalty for being immoral, it is often ambiguous whether a highly public display of moral behavior stems purely from a selfless moral motivation or instead from a more egoistic motivation, such as a concern with one's own moral reputation.

Third, and relatedly, self-reports of one's own moral character are not necessarily reliable. For one thing, direct reports of one's own good character, or of behaviors that illustrate one's good character, are likely seen as non-diagnostic for the same reason that some public moral performances are non-diagnostic—there is a similar ambiguity as to their underlying motivation. Part of this non-diagnosticity stems from the fact that direct reports of

[8] There are notable exceptions. For instance, traits such as temperance may need to be frequently exercised. However, while lapses of temperance may be quite visible, positive displays of temperance may be considerably less visible (it is not always easy to detect that someone is restraining themselves), thereby complicating the process of inferring temperance from behavior.

one's own good character can easily be manipulated in self-serving ways, in contrast to indirect reports from second and third parties, which observers tend to afford greater credence (Walther, Van Der Heide, Hamel, & Shulman, 2009). Additionally, reporting directly about one's own moral character can come across as boastful and narcissistic, thereby undercutting the perceived ethical motivation for the behavior in question (Berman, Levine, Barasch, & Small, 2015). Moral boasting may also be perceived as defensive and insincere (the "lady doth protest too much"). Finally, moral boasting may also be judged by the recipient as a kind of moral reproach, that is, as a kind of implicit moral judgment of them (see e.g., Minson & Monin, 2012). Accordingly, astute social actors may tend to guard against reporting on their own morally good deeds, thus making such reports relatively infrequent. In sum, a multitude of factors diminish both the likelihood and diagnosticity of direct first-person reports of one's own moral character, thereby further complicating the task of drawing character inferences about others in daily life.

Together, these three features—scarce opportunities for strictly obligatory moral behavior, the ambiguity of highly public moral behavior, and the ambiguity and social complexity of reporting one's own moral behavior—combine to make inferences about other people's moral character in ordinary life highly complex. Because we know so little about how these naturalistic inferences of character actually take place, we think that they warrant further scrutiny.

WHAT IS THE PERCEIVED STRUCTURE OF MORAL CHARACTER?

A third unresolved issue, and perhaps the most perplexing, pertains to the underlying structure of perceived moral character. Two distinct questions emerge here. First, what is it that unifies moral character traits? And second, what divides them?

With respect to the first question, we have argued that moral traits are separate from other sorts of traits, including warmth traits. And our results show that people are clearly capable of identifying the degree to which a trait is diagnostic of morality, or not, which suggests that there is some general feature that unifies them. But we have not precisely identified what this feature is (if anything). What is it that makes something a moral trait in the eyes of ordinary observers? Is it that such traits reflect a person's deeper helpful or harmful intentions toward others (see e.g., Abele & Wojciszke, 2007; Wojciszke, Bazinska, & Jaworski, 1998; Wojciszke, Dowhyluk, & Jaworski, 1998; see also Cottrell et al., 2007; Cuddy et al., 2008; Fiske et al., 2007; Leach et al., 2007)? Is it that such traits epitomize selfless concern (or the lack thereof) for others' welfare (e.g., Schopenhauer, 1995/1841)? Or is it that such traits embody a concern (or the lack thereof) with an adherence to a particular set of moral rules (e.g., Kant,

1981). Or does no single idea capture what is fundamental to moral character traits? We are not sure, but we think this is an important avenue for future research.

However, even though there may be some common thread that underlies moral character traits, there are also clear and important differences between the moral virtues as well. Kindness is clearly not the same thing as honesty, and both are, in turn, different from loyalty. Indeed, this is brought into stark relief when moral virtues collide, such as when it would be unkind to tell the truth, or when it would be unjust to be loyal. More generally, the list of enumerable moral traits is perhaps limitless (see e.g., Russell, 2009). The question we are concerned with is how we should think about the structure of these moral virtues in person perception. Is there some dimensional structure to the moral virtues, so that some are seen as closely related enough that they reduce to the same underlying dimension? Is there a set of distinct, foundational, moral virtues? And, separately, is there any hierarchical organization of the virtues, such that some are more important than others, or that some are dependent on others?

Several existing theories within psychology maintain a commitment to the idea that there is a dimensional structure to the moral virtues. Although the list of discrete moral virtues that could be enumerated is long, and potentially limitless, these psychological theories argue that the virtues devolve into a small handful of primary dimensions. For instance, Jonathan Haidt, Jesse Graham, Craig Joseph, Brian Nosek, and colleagues' moral foundations theory maintains that five core moral virtues correspond to each of the five different moral "foundations" or "intuitions." (see also, Haidt & Joseph, 2004, 2007). For instance, the virtues corresponding to the five foundations outlined in Graham, Haidt, and Nosek's (2009) theory (with the corresponding moral foundation in parentheses) are thought to be compassion (harm), justice (fairness), loyalty (in-group), respectfulness (authority), and chastity/temperance (purity). By contrast, Lawrence Walker and Karl Hennig have focused specifically on moral personality, dividing it into three central types: just, brave, and caring (Walker & Hennig, 2004), a scheme which only partially overlaps with the five foundations scheme. Another, still different taxonomy is Peterson and Seligman's (2004), which argues for the following six distinct virtues: wisdom and knowledge, courage, humanity, justice, temperance, and transcendence, each of which further divides into subvirtues.

Several philosophers have also devised moral theories that crystallize the core elements of moral character. Bernard Gert, for instance, conceptualizes morality in terms of two broad categories of moral rules, harm and trust, and in terms of a single moral ideal, which is to lessen the amount of harm or evil in the world (Gert, 2004, 2005). According to Gert, a single trait, the *absence of cruelty* encourages the avoidance of harm, while several more-specific virtues correspond to the rules governing trust (see, e.g., Gert, 2009): truthfulness, trustworthiness,

justice, law-abidingness (obeying the law), and conscientiousness. Finally, kindness is the virtue that promotes following the moral ideal of lessening harm.[9]

W. D. Ross's work also yields a taxonomy of different ethical virtues (Ross, 1930, 1939). Although there is some debate about the exact number of *prima facie* (or perhaps more accurately, 'pro tanto,' see Kagan, 1989) duties Ross posited, it is somewhere around seven, or possibly eight: fidelity (including both promise-keeping and honesty), reparation (righting one's previous wrongs), gratitude, non-malificence (not injuring or harming others directly), beneficence (doing good to others, or preventing harm to them from others),[10] self-improvement (particularly in relation to moral matters), and justice (see e.g., Garrett, 2004; Simpson, 2012; Skelton, 2012). Each of these duties seems to correspond directly to a specific moral virtue, with the possible exception of self-improvement.

Aristotle was another thinker whose ethical theorizing in the *Nichomachean Ethics* yields a list of approximately 10 to 11 core ethical virtues (see e.g., Bartlett & Collins, 2011, pp. 303–304), as was Benjamin Franklin, who adumbrated a set of 13 core moral virtues.

Thus, while it is far from exhaustive, this brief survey illustrates that questions about the structure of moral character, including which moral character traits are fundamental to the ethical life, has exercised philosophers as well as psychologists. It also illustrates that philosophers, like psychologists, disagree regarding which moral virtues are most fundamental. (They also disagree regarding whether any such hierarchy of traits exists in the first place.)

Philosophers have been concerned with normative questions about virtue and duty—which virtues and duties, if any, are in fact fundamental—whereas psychologists have wondered about which virtues are seen as fundamental by ordinary individuals, which is our primary interest. However, the methods employed in the psychological literature thus far are limited in that they employ highly top-down approaches, starting with a priori theoretical commitments to a particular set of virtues, and then examining judgments made in relation to those virtues. While such methods illuminate the particular virtues in focus, they are inadequate to address the question of which virtues, or which dimensions of virtue, are seen as fundamental in the first place.

Accordingly, we have begun to take preliminary steps in order to answer this question (Goodwin & Piazza, 2015). In one study, we asked participants to rate each of the 170 traits used in our earlier database (Goodwin et al., 2014,

[9] Gert also considers virtues such as courage, prudence, and temperance, which he regards as "personal virtues," rather than moral virtues per se. He regards virtues of this sort as related to the moral virtues, in that they are necessary in order to express moral virtues, but not moral virtues in and of themselves (see Gert, 2004, pp. 77–78).

[10] These two aspects of beneficence, positively doing good to others and preventing harm to others from causes other than oneself, have sometimes been separated (see, e.g., Garrett, 2004).

Study 1) on a set of new questions. Each trait was rated on how useful it would be for judging whether another person possessed each of 13 broad candidate dimensions of virtue: just/fair, respectful/polite, trustworthy/dependable, kind/compassionate, honest/sincere, humble/modest, obeys/respects authority, loyal/faithful, self-controlled/disciplined, strong-willed/determined, hard-working/productive, courageous/brave, and wise. We selected these broad dimensions because they are semantically distinct, and because they each seem to represent an aspect of moral virtue that could be fundamental. We then had two sorts of ratings for each of the 170 traits—the extent to which they indexed (were useful for judging) each of these 13 specific candidate dimensions of character, and the extent to which they indexed (were useful for judging) character more broadly (these latter ratings came from our earlier Study 1; see Goodwin et al., 2014). We were particularly interested in which of the 13 more specific virtue-relevance ratings would most strongly predict the broader character-relevance ratings, which we assessed via regression analysis. Doing this allowed us to answer the question, which dimensions of virtue are seen as strongly and independently related to the notion of having a good moral character? Our aim, therefore, was to narrow down the initial set of 13 virtue dimensions to a somewhat reduced set of virtue dimensions that independently predict relevance to moral character. If it turned out that each of the 13 dimensions independently predicted relevance to moral character, that would also be revealing.

This analysis showed that four dimensions independently predicted relevance to moral character: kindness/compassion, honesty/truthfulness, justice/fairness, and self-control/discipline. The first three dimensions showed particularly strong predictive links to moral character. Notably, the moral character traits revealed to be most fundamental by this study only partially overlap with past psychological investigations that have employed top-down approaches (e.g., Graham et al., 2009; Haidt & Joseph, 2004, 2007; Walker & Hennig, 2004). And while this is only one initial study, we think it is an important first step in trying to establish which virtues are psychologically fundamental in an empirically driven, bottom-up way.

Another study we recently conducted aimed to provide some sense of the implicit ranking that ordinary individuals attribute to different moral virtues (Goodwin & Piazza, 2015). We presented participants with the same list of 13 candidate dimensions of virtue used in the study just described, and asked them to rank order them in terms of their "importance for determining whether a person has moral character." This revealed that honesty/sincerity was regarded as the most important dimension, significantly more so than any other trait dimension. Next underneath honesty/sincerity were kindness/compassion, justice/fairness, and trustworthiness/dependability, which were ranked about equally, and higher than all the other candidate moral virtues. Thus, this procedure produced results that were reasonably consistent with the

results of the initial regression study, with the exception being that trustworthiness/dependability rose in prominence relative to self-control/discipline. Nonetheless, both studies converge on the conclusion that kindness/compassion, honesty/sincerity, and justness/fairness are perceived as among the most fundamental moral virtues. These initial results need to be supplemented with further convergent evidence. But even at this early stage, they represent promising new lines of evidence as to the perceived structure of moral character.

In a separate line of research (Piazza, Goodwin, Rozin, & Royzman, 2014), we have also found that people perceive virtues of kindness, honesty, trustworthiness, and fairness to be *unconditionally* good, such that when a person with evil values (e.g., a white supremacist or terrorist) possesses one or other of these traits, that person is judged to be a morally better person than if they did not possess such a trait (we refer to these traits as *core goodness traits*; see Piazza et al., 2014). Such core goodness traits may be contrasted with what we refer to as *commitment traits*, such as discipline, determination, and dedication, which are only conditionally good. That is, while these commitment traits are seen as highly relevant to moral character, and as enhancing positive evaluations of good or neutral agents (thereby being "default positive"), they also amplify the *immorality* of persons who endorse evil values (see Piazza et al., 2014). We therefore speculate that whether or not a trait is perceived as unconditionally good may predict how fundamental to moral character it is seen as being.

Conclusion

Kupperman (1991) noted over 20 years ago that, in both psychology and philosophy, moral character had received little attention in comparison with the attention paid to specific moral dilemmas. Of course, within philosophy, character had received considerable attention in the context of virtue ethics (e.g., Anscombe, 1958; MacIntyre, 1981). But Kupperman thought that even the virtue ethicists did not sufficiently emphasize the relation between the virtues in constituting the overall character of a person (see Kupperman, 1991, chap. 5). Kupperman's analysis alerted us to a similar marginalization of character within psychology, and it stimulated our initial work. Other researchers have likely had different motivations, but have similarly been drawn to a new perspective that emphasizes the role of moral character in both person perception and personality.

Moral character is now making a comeback (Pizarro & Tannenbaum, 2011). As the other chapters in this volume reveal, a direct focus on moral character traits reveals many new insights into people's behavior. Moreover, the study of moral psychology, and moral judgment in particular, has recently been reinvigorated by a renewed focus on the role of moral character in moral judgment. A

dominant theme of this emerging literature is that ordinary moral judgments take into account not just discrete moral or immoral actions, but also information about the underlying character of the perpetrators of those actions (see, e.g., Alicke, 1992, 2000; Alicke, Buckingham, Zell, & Davis, 2008; Nadler & McDonnell, 2011; Pizarro & Tannenbaum, 2011; Tannenbaum, Uhlmann, & Diermeier, 2011). In other words, judgments of character make up an important part of the overall process of moral appraisal.

Our focus is somewhat different, in that we have concentrated on the role of moral character in person perception and cognition (see also Brambilla & Leach, 2014). But here too, we believe that a renewed emphasis on the moral dimensions of personality has begun to reveal new insights into human social cognition, while giving rise to new and important questions that remain to be answered.[11]

References

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology*, *47*, 935–948.

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*, 1063–1065.

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*, 1202–1217.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763.

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574.

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin*, *34*, 1371–1381.

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt.

Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy*, *33*, 1–19.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 1230–1240.

[11] This chapter was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. The authors thank Christian Miller and Mike Furr for their valuable comments on earlier drafts of this chapter, and the audiences at the Character Project Conferences (Wake Forest University, 2012 and 2013) for their stimulating feedback.

Asch, S. E. (1952). *Social psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*, 327–353.

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, *87*, 707–721.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., de Vries, R. E., Di Blas, L., . . . De Raad, B. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycho-lexical studies in seven languages. *Journal of Personality and Social Psychology*, *86*, 356–366.

Bartlett, R. C., & Collins, S. D. (2011). *Aristotle's Nichomachean Ethics: A new translation*. Chicago: University of Chicago Press.

Bauman, C. W., & Skitka, L. J. (2012). Corporate social responsibility as a source of employee satisfaction. *Research in Organizational Behavior*, *32*, 63–86.

Baumeister, R. F., Bratslavsky, E., & Finkenauer, C. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.

Berman, J. Z., Levine, E. E., Barasch, A., & Small, D. A. (2015). The braggart's dilemma: On the social rewards and penalties of advertising prosocial behavior. *Journal of Marketing Research*, *LII*, 90–104.

Bermant, G., Talwar, C., & Rozin, P. (2011). To celebrate positive psychology and extend its horizons. In K. M. Sheldon, T. B. Kashdan, & M. F. Steger (Eds.), *Designing positive psychology: Taking stock and moving forward* (pp. 430–438). Oxford: Oxford University Press.

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*, 397–408.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*, 135–143.

Brambilla, M., Rusconi, P., Sacchi, S., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*, 149–166.

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, *49*, 811–821.

Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization*, *43*, 57–73.

Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different desired characteristics. *Journal of Personality and Social Psychology*, *92*, 208–231.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149.

Darley, J. M., & Batson, D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*, 100–108.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*, 629–636.

DeSteno, D., & Valdesolo, P. (2011). *Out of character.* New York, NY: Crown Archetype.

Doris, J. M. (2002). *Lack of character*. Cambridge, UK: Cambridge University Press.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*, 77–83.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902.

Garrett, J. (2004). A simple and usable (though incomplete) ethical theory based on the ethics of W. D. Ross. Retrieved from http://people.wku.edu/jan.garrett/ethics/rossethc.htm

Gert, B. (2005). *Morality: Its nature and justification* (Rev. ed.). Oxford: Oxford University Press.

Gert, B. (2004). *Common morality: Deciding what to do*. Oxford: Oxford University Press.

Gert, B. (2009). Toward a common morality. Retrieved from http://www.youtube.com/watch?v=enVFjAUTfI8

Goodwin, G. P., & Piazza, J. (2015). *What is the perceived structure of moral character?* Manuscript in preparation.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046.

Haidt, J., & Joseph, C. (2004). How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*, 55–66.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence & S. Stich (Eds.) *The Innate Mind,* (Vol. 3, pp. 367–391). New York: Oxford University Press.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character*. New York: Macmillan.

Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*, *21*, 384–388.

Kagan, S. (1989). *The limits of morality*. Oxford: Clarendon Press.

Kamtekar, R. (2004). Situationism and virtue ethics on the content of our character. *Ethics*, *114*, 458–491.

Kant, I. (1981). *Lectures on ethics.* Indianapolis, IN: Hackett Publishing Company.

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality*, *18*, 431–439.

Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology*, *48*, 77–85.

Kinder, D. R., Peters, M. D., Abelson, R. P., & Fiske, S. T. (1980). Presidential prototypes. *Political Behavior*, *2*, 315–337.

Kupperman, J. (1991). *Character.* New York: Oxford University Press.

Kupperman, J. (2001). The indispensability of character. *Philosophy*, *76*, 239–250.

Landy, J. F., Piazza, J., & Goodwin, G. P. (2015). *Morality, sociability, and competence are distinct dimensions of social cognition.* Manuscript under submission.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*, 234–249.

Lin, M. H., Kwan, V. S. Y., Cheung, A., & Fiske, S. T. (2005). Stereotype Content Model explains prejudice for an envied outgroup: Scale of anti-Asian American Stereotypes. *Personality and Social Psychology Bulletin*, *31*, 34–47.

MacIntyre, A. (1981). *After virtue*. Notre Dame, IN: University of Notre Dame Press.

Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*, 661–684.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371–378.

Miller, C. B. (2003). Social psychology and virtue ethics. *The Journal of Ethics*, *7*, 365–392.

Miller, C. B. (2009). Empathy, social psychology, and global helping traits. *Philosophical Studies*, *142*, 247–275

Miller, C. B. (2010). Character traits, social psychology and impediments to helping behavior. *Journal of Ethics and Social Philosophy*, *5*, 1–36.

Miller, C. B. (2014). *Character and moral psychology*. Oxford: Oxford University Press.

Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, *3*, 200–207.

Nadler, J., & McDonnell, M. H. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review*, *97*, 255.

Nicholson, I. (1998). Gordon Allport, character, and the "culture of personality," 1897-1937. *History of Psychology*, *1*, 52–68.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification.* Washington, DC: APA Press and Oxford University Press.

Piazza, J., Goodwin, G. P., & Rozin, P. (2015). *Moral character evaluations are based on a person's weakest link* Manuscript in preparation.

Piazza, J., Goodwin, G., Rozin, P., & Royzman, E. (2014). When a virtue is not a virtue: Conditional virtues in moral evaluation. *Social Cognition*, *32*, 528–558.

Pizarro, D. A., & Tannebaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington DC: American Psychological Association.

Reeder, G. D., & Coovert, M. D. (1986). Revising an impression of morality. *SocialCognition*, *4*, 1–17.

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, *44*, 736–745.

Regan, D. T., Williams, M., & Sparling, S. (1972). Voluntary expiation of guilt: A field experiment. *Journal of Personality and Social Psychology*, *24*, 42–45.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–294.

Rosenberg, S., & Olshan, K. (1970). Evaluative and descriptive aspects in personality perception. *Journal of Personality and Social Psychology*, *16*, 619–626.

Ross, W. D. (1930). *The right and the good*. Oxford: Oxford University Press.

Ross, W. D. (1939). *Foundations of ethics*. Oxford: Oxford University Press.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.

Russell, D. C. (2009). *Practical intelligence and the virtues*. Oxford: Oxford University Press.

Sabini, J. & Silver, M. (2005). Lack of character? Situationism critiqued. *Ethics*, *115*, 535–562.

Schopenhauer, A. (1995). On the basis of morality (E. F. J. Payne, Trans.) Indianapolis, IN: Hackett. (Original work published 1841.)

Schopenhauer, A. (2009). *The two fundamental problems in ethics* (C. Janaway, Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1841).

Seligman, M. E. P. (2002). *Authentic happiness*. New York: Basic Books.

Simpson, D. L. (2012). William David Ross (1877–1971). In *The internet encyclopedia of philosophy*. Retrieved from: http://www.iep.utm.edu/ross-wd/

Singer, P. (1979). *Practical ethics*. Cambridge, UK: Cambridge University Press.

Singer, P. (1995). *How are we to live: Ethics in an age of self-interest*. Amherst, NY: Prometheus Books.

Singer, P. (2009). *The life you can save*. New York, NY: Random House.

Skelton, A. (2012). William David Ross. In E. N. Zalta (Ed.,), *The Stanfordencyclopedia of philosophy*. Retrieved from: http://plato.stanford.edu/archives/sum2012/entries/william-david-ross/

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*, 689–699.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*, 159–171.

Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, *47*, 1249–1254.

Trafimow, D., & Trafimow, S. (1999). Mapping perfect and imperfect duties onto hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin*, *25*, 687–697.

van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision making. *Cognition*, *108*, 796–803.

Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology*, *86*, 629–647.

Walther, J. B., Van Der Heide, B., Hamel, L., & Shulman, H. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using Facebook. *Communication Research*, *36*, 229–253.

Warm. The Free Dictionary online (n.d.). Retrieved from http://www.thefreedictionary.com/warm.

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*(October 24), 606–607.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-Ms exposure to a face. *Psychological Science*, *17*, 592–598.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1251–1263.

Wojciszke, B., Dowhyluk, M., & Jaworski, M. (1998). Moral and competence-related traits: How do they differ? *Polish Psychological Bulletin*, *29*, 283–294.

PART } III

The Existence and Nature of Character

5 }

Moving Character beyond the Person-Situation Debate: The Stable and Dynamic Nature of Virtues in Everyday Life

Wiebke Bleidorn

"You can't trust him, he's got weak morals and ethics—and he's done a heck of a good job." This is how ABC News summarized the public consensus on former US president Bill Clinton when he left office in 2001 (Langer, 2001). Despite his sexual misadventures and his impeachment by Congress, the vast majority of Americans (65%) approved of the way Clinton had handled his job as president. In fact, his approval rating still stands out as the best end-of-career rating of any postwar president. More than two-thirds of Americans endorsed his performance on the economy, saw him as a strong leader, and valued his work on race relations. At the same time, his personal ratings hit the rocks; 76% of Americans rated him as not honest and trustworthy; 77% said he lacks moral and ethical standards, and only 44% percent viewed him favorably "as a person" (cf. Langer, 2001).

Referring to Clinton's performance as a president and his social engagement since he left office, the media frequently described him as talented, passionate, hard-working, competent, charming, and optimistic. However, the same magazines that raved about Clinton's qualities on the job dogged him for his equally obvious moral weaknesses. Indeed, Clinton admitted that he had lied about his extramarital relationship with Monica Lewinsky to his wife, his daughter, Congress, and the American public. In his autobiography, he later explained, "Since 1991 I had been called a liar for everything under the sun, when in fact I had been honest in my public life and financial affairs, as all the investigations would show. Now I was misleading everyone about my personal failings. I was embarrassed and I wanted to keep it from my wife and daughter" (Clinton, 2005, p. 775).

How shall we interpret Clinton's conduct? Did he lie because dishonesty is an innate quality of his character, or did he decide not to tell the truth to protect his family and to prevent an apparently inevitable media battle? Does his dishonest behavior indicate a lack of virtue? Or does his candid performance on the job better reflect his character? Is it possible to be both honest at one time and not at another? Can a "good character" inoculate against moral failures? Or does the apparent inconsistency in virtuous behavior instead prove that there is no such thing as character? Is it merely situations that lead to both people's most noble acts and their worst moral failures?

These questions have not only been discussed with regard to the former president's character (e.g., Goren, 2002). They lie at the heart of an ongoing debate about the nature and very existence of human character. Philosophers (e.g., Alfano, 2013; Doris, 2002) and psychologists (e.g., Darley & Batson, 1973) have wrestled with the questions of whether and to what degree a person's moral conduct might reflect his character or, instead, the situation he is acting in. A still highly controversial point at issue concerns the implications of behavioral (in-)consistency. Why are we upright and prudent in some situations but may be cheaters and rogues in others? In view of the apparent inconsistency and unpredictability of virtuous behavior, isn't it naïve to believe in stable virtues and character traits? The consistency issue and the person-situation debate (Kenrick & Funder, 1988) have challenged the very existence of character traits, and although the debate has been declared over more than once, its aftermath is still noticeable in scientific psychology and beyond (Epstein & O' Brien, 1985; Fleeson, 2004; Roberts & Caspi, 2001).

The present chapter is dedicated to moving character psychology beyond the person-situation debate by endorsing both the consistent and dynamic nature of character traits and virtuous behavior. Based on recent theoretical and empirical work in the field of personality psychology, it proposes (1) that there is more than one important type of behavioral consistency relevant to character (e.g., Fleeson & Noftle, 2008b), (2) that relatively stable character traits can be described as density distributions of dynamic and changing virtue states, and (3) that people are able to flexibly adjust their character-relevant behavior to the demands of their current role context.

Resolving the consistency issue and moving character beyond the person-situation debate are important for at least three reasons. First, resolving the consistency issue will help to end the long-lasting impasse in research on character in personality psychology. Second, personality psychologists will move beyond the question of whether character exists to the important question of how it works and develops. Third, the resolution of the consistency issue in character psychology may fertilize new theorizing and research, not only in psychology but also in other disciplines (cf., Fleeson & Noftle, 2008a; Miller, 2013).

After a brief historical review of the rise and fall of character in scientific psychology, I will present the theoretical and empirical background for the aforementioned hypotheses in more detail, discuss recent findings from my own research, and point out open issues that need to be addressed by future studies.

The Rise and Fall of Character in Scientific Psychology

THE HEYDAY

Research on character, virtues, and moral behavior flourished during the early years of scientific psychology. In the field of personality psychology, no fewer than 1,908 articles and 483 books on character were published between 1890 and 1940 (http://www.ebscohost.com/academic/psycinfo). Scholars from all kinds of schools and backgrounds were interested in the nature, meaning, and consequences of character and moral behavior. Thus, the early psychological writings on character covered an extremely broad scope of topics ranging from, for example, the "physique of character" (Kretschmer, 1925) to the benefits of and problems with character education (Hartshorne & May, 1930) and the use of "slang as an indicator of character" (p. 245; Schwesinger, 1926). Among the most cited articles is a review article by Gordon W. Allport (1921)[1] in which he summarized the state of research on personality and character at the time:

> The enticing topic of speculation, human nature, has been rephrased for scientific study as a problem of individual nature. On every hand we see that interest is becoming focused upon the study of man with the hope that by this method the problems of the mechanisms, motives, and modifications of behavior peculiar to the individual may at length be solved. This tendency is manifested by the rapid increase in the number of studies of human personality and character (p. 441).

THE TROUBLES WITH CONSISTENCY

While most scholars agreed that traits offer an elegant and parsimonious way to capture an individual's conduct, the question emerged of how forceful and influential traits can be in view of the apparent power of the context

[1] Already in this early work, as well as in his later writings, Allport (1937) stressed the importance of distinguishing personality and character traits; and in his still influential textbook he noted, "Character is personality evaluated and personality is character devaluated" (p. 52). That is, both character and personality traits can be broadly defined as relatively stable dispositions to behave in a particular way over a range of situations. However, in contrast to personality traits, character traits always appear to have an evaluative dimension.

and situation an individual is acting in. According to Allport (1927), the situational dependency of behavior is particularly problematic with regard to character traits, "The same behavior psychologically speaking may be moral in one locality, immoral in another, moral at one period of time, immoral at another" (p. 285).

The crisis probably began with Hartshorne's and May's (1928) still famous "Character Education Inquiry," a then seminal empirical investigation of honest and deceptive conduct in schoolchildren. In several studies, Hartshorne and May found that even across relatively similar situations, honest and dishonest behavior were shown inconsistently. For example, some kids were honest with their peers, but cheated in classroom situations or lied to their parents. From these studies, Hartshorne and May concluded that character traits are not robust or consistent but rather specific functions of life situations. Around the same time, Newcomb (1929) published a study of extraverted and introverted behavioral patterns in problem boys. He found "practically no trait consistency" and agreed with Hartshorne and May that trait-relevant behaviors seem to be highly situation specific.

Findings like these have shaped the term *situationism*. This research perspective has stressed the influence of the situation over that of the person in predicting an individual's behavior (for reviews, see, e.g., Doris, 2002; Epstein & O'Brien, 1985; Miller, 2014). The situationist's perspective challenged not only the character approach but trait theories in general.

THE DOWNFALL

The eventual downfall of character traits came in 1968, precipitated by the publication of the now classic monograph, *Personality and Assessment* by Walter Mischel. The book touched upon the problem in trait assessment that had already been identified by Allport back in 1937. Mischel argued that traditional conceptions of traits are based on the theoretical prediction of behavioral consistency, which, however, cannot be demonstrated empirically. In his review, he listed a multitude of studies that had failed to support the fundamental proposition that an individual's trait-relevant behavior is highly consistent across diverse situations. Instead, Mischel's analyses again emphasized that an individual's behavior was highly dependent upon situational cues and contextual circumstances.

Although the consistency issue had been already discussed earlier, Mischel's book came in like a wrecking ball and led to a paradigm crisis in personality psychology that changed the agenda of the field for decades. In the spirit of that time, several sensational and still highly controversial studies were published which questioned the existence of traits. For example, in one of the most cited experimental studies of that time, Darley and Batson (1973) examined helping behavior in students at the Princeton Theological Seminary who were

preparing for the ministry. In this study, the religious views and character traits of the students did not predict which students helped a person in need on their way to an appointment. Rather, the best predictor of students' willingness to help was whether they were in a hurry or had some time to spare until their appointment. This study and many other "situationist" experiments (e.g., Milgram, 1974; Haney, Banks, & Zimbardo, 1973) added fuel to the fire, causing many psychologists to question and doubt whether personality or character exist at all.

The person-situation debate has plagued personality psychology over the ensuing decades. Proponents of the "person camp" have argued that, because behavior is determined in large part by a person's traits, individuals will act similarly to themselves much of the time, except for some reasonable adaptation to changing circumstances. In response to the situationist's critique, a large body of research has been done to provide empirical evidence for the existence and usefulness of trait concepts. And today, there actually is considerable consensus that individual differences in broad personality traits are relatively stable over time (e.g., Roberts & DelVecchio, 2000), substantially influenced by genetic and environmental factors (e.g., Bleidorn et al., 2009; Bleidorn, Kandler, & Caspi, 2014), and important predictors of relevant life outcomes, such as occupational success or even mortality (Ozer & Benet-Martinez, 2006; Roberts et al., 2007). However, the situationist's argument remained that, because the immediate situation is the primary determinant of behavior, individuals will act very differently on different occasions. This position was strengthened by the burgeoning field of social psychology, and in numerous studies, scholars demonstrated powerful and pervasive effects of situations on people's behavior (for a review, see Ross & Nisbett, 1991).

The Postwar Period of the Person-Situation Debate: Modern Perspectives on the Consistency Issue

After a long "stalemate" (Fleeson & Noftle, 2008a, p. 1671), the person-situation debate has come to an end (e.g., Donnellan, Lucas, & Fleeson, 2009), and today, most scholars agree that both the person and the situation play an important part in influencing an individual's behavior. That is, the debate has moved from pitting person and situational factors against one another toward examining how each can inform the other about the ways in which traits might affect an individual's behavior in a particular situation (Bleidorn, 2009; Funder, 2008; Heller, Komar, & Lee, 2007). Hence, the original question of whether there is behavioral consistency at all has been transformed into the questions of how and under which conditions trait-relevant behavior might show a certain degree of consistency (Fleeson & Noftle, 2008b). An important step toward this resolution was the realization that there is not just

one type of behavioral consistency but multiple types of consistency, and that traits may be highly consistent in some ways but not in others (Fleeson & Noftle, 2008b, 2008a).

THE DENSITY-DISTRIBUTION APPROACH TO PERSONALITY

Based on the notion that there is more than one type of consistency and building upon the earlier work of Epstein (1979) and Mischel (1968), among others, Fleeson (2001) proposed a theoretical model that aimed to integrate both the consistent and inconsistent parts of personality traits. In his density-distribution approach to personality, Fleeson proposed that traits should not be conceived as single ways of acting, but rather as entire frequency distributions of their corresponding *states* (see Figure 5.1).

Personality states were defined as reflecting an individual's personality in a given moment rather than in general—the more the person manifests a certain personality trait, the higher the level of the corresponding state. For example, the more a person manifests behaviors that reflect trait extraversion, the higher her momentary level of state extraversion is. States can be measured in the same way as traits, using the same content (e.g., adjectives) and the same dimensions, but they describe how much a person manifests those traits *in a given moment* rather than in general (Bleidorn, 2009; Fleeson, 2001; McCabe & Fleeson, 2012). Just as an individual can be described by means of his or

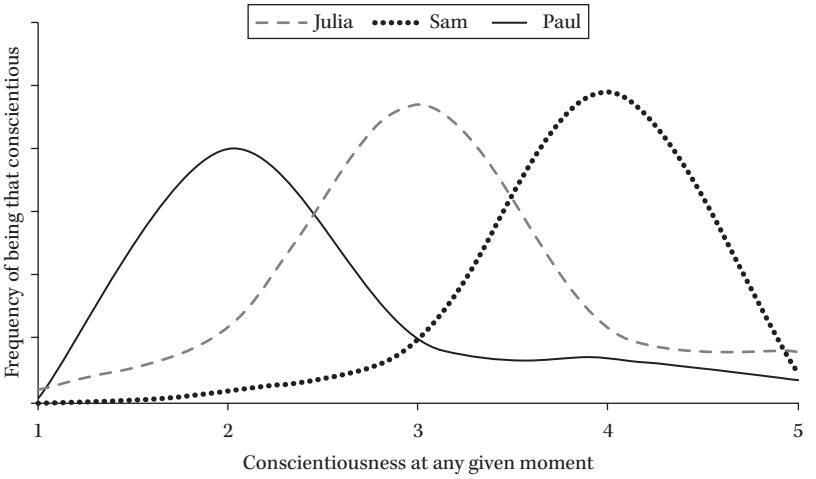

FIGURE 5.1 **Traits as Density Distributions.** The three distributions show the conscientiousness levels of three hypothetical individuals across a week of behavior. During this week, all three individuals show all levels of conscientiousness more or less often. That is, they sometimes act very conscientiously and sometimes not at all conscientiously, suggesting that behavior is inconsistent. At the same time, each individual's distribution is anchored around a mean-level of conscientiousness that is most frequently shown by an individual

her general level in agreeableness, an individual's momentary behavior can similarly be described as being more or less agreeable. That is, the approach does not focus on specific behavioral acts (e.g., lying to Congress), but is an assessment of how the person is behaving with regard to trait expression (e.g., honesty).

Over time and across situations, each individual's various momentary states will form a distribution indicating the frequency with which the individual is at each level of the state. Fleeson (2001) proposed that although single behaviors are less predictable, the average level of these distributions should be highly predictable and relatively stable over time. Furthermore, not only the average levels, but also the size of the distributions (i.e., the variability itself) should prove to be a relatively stable and predictable individual differences characteristic.

In summary, the density distribution approach conceives personality traits as entire density distributions rather than as only one aspect of the distributions. Hence, this approach can account for both the low cross-situational consistency of single behaviors and the relatively high consistency of averaged behavior (Fleeson, 2001, 2007).

EMPIRICAL FINDINGS

A critical examination of the density-distribution approach to personality requires different research designs and methodology than standard personality questionnaires or single experiments. A meaningful test of the hypothesis that traits can be described as density distributions of personality states necessitates observing people as they live their lives and collecting a large number of behavioral assessments across time and situations.

To this end, scholars have begun to use modern data-collection technology and experience-sampling designs (for a critical review of this methodology, see Scollon, Kim-Prieto, & Diener, 2003). In a typical experience-sampling study, participants carry some kind of data-collection device (e.g., personal data assistant, microcomputer, or smartphone) for a couple of days or even weeks, which they use to record their momentary behavior multiple times per day (e.g., Bleidorn, 2009; Fleeson, 2001, 2007; McCabe & Fleeson, 2012). Over this period, each participant provides a multitude of momentary behavioral reports, which allow looking at single behavioral states, the average levels of these states, and also the variability in state expression. The accumulation of an individual's personality states over time and situations forms a distribution displaying the frequency with which the individual manifested the state at each level along the dimension's scale.

Using this experience-sampling methodology, several studies have shown that individuals express nearly all levels of all personality traits in their everyday behavior (e.g., Bleidorn, 2009; Fleeson, 2001, 2007). For example, individuals

who were on average relatively conscientious still showed a wide range of state levels ranging from extremely conscientious to highly unconscientious states. That is, the typical density distribution was wide and best described by a personal mean-level plus variable states around this mean.

However, this high degree of variability did not imply that individuals did not differ from each other in the average state levels they expressed. Rather, individuals' distributions were typically centered around their personal means, and individuals differed reliably from each other in these personal means. Furthermore, these personal mean-levels were highly stable from week to week, with correlations ranging around r = .80 to r = .90 (Fleeson, 2001; Bleidorn, 2009). The density-distribution approach thus acknowledges both the high consistency of averaged behaviors and the cross-situational inconsistency of single behaviors.

So far, the density-distribution approach has proved to be a useful model in describing broad personality traits such as the Big Five (i.e., neuroticism, extraversion, openness, agreeableness, and conscientiousness; for a review, see John, Naumann, & Soto, 2008). Specifically, several studies have shown that this model provides a unique perspective on personality that allows us to examine the conditions and consequences of intra-individual variability in personality-relevant behavior while taking into account the stable nature of the potentially underlying trait dimensions. For example, McNiel and Fleeson (2006) found that when generally introverted individuals acted extraverted, their state extraversion had affective consequences. Specifically, their state extraversion was linked to momentary positive affect, suggesting that even introverted individuals were in a better mood when acting extraverted. Furthermore, people seem to change their personality states in response to situational cues and role contexts (Bleidorn, 2009; Fleeson, 2007), possibly in order to accomplish personal goals (Heller et al., 2007; McCabe & Fleeson, 2012). Thereby, they report that these personality states feel true to themselves and are not just hollow acts, faking, or insincere forced behaviors (Fleeson & Wilt, 2010).

In summary, the density-distribution approach to personality informs the questions of whether, when, and how traits may be consistent or not. Moreover, it offers a new way of looking at traits, a perspective that acknowledges both the consistent and the dynamic aspect of traits. This new perspective on traits has fruitful consequences for both theory and research on personality structure and functioning (for a review, see, Fleeson, 2012).

A Density-Distribution Approach to Character

Can Fleeson's density-distribution approach be applied to character traits and virtuous behavior? There may be reasons to believe that the density-distribution approach cannot be easily translated to character traits. Specifically, according

to Peterson and Seligman (2004), a distinctive feature of character traits is their inherent positivity: that is, the presumption that character traits are morally valued in their own right. It could thus be argued that character traits are always good in all situations. A person with high moral standards should therefore consistently manifest his or her character traits across situations and time. In this case, we would expect to find character differences among individuals, whereas momentary differences within individuals should be negligible.

Alternatively, it could be argued that character traits, despite their inherent positivity, are still contingent on contextual features and situational attributes. Put differently, certain character traits might be more relevant in some situations as compared to others, and there might be times at which showing a lower level of a character trait might actually have beneficial effects for the individual.

So, the question remains whether it is useful to conceive broad virtue domains like wisdom, courage, or humanity as density distributions of virtue states. Do people regularly switch from acting wisely to naively, from bravely to cowardly, from telling lies to being honest, just as they switch from being extraverted to introverted? Are people as variable in their momentary expression of their virtues as they are with regard to their personality traits or even affect? If so, are individuals still characterized by unique, predictable, and stable distributions of virtue states?

These questions lay at the heart of the project *Loving Mum and Tough Career Girl—How Do Role Settings Affect Our Personality and Character Traits?*, which aimed to take a fresh look at the consistency of character traits from the perspective of contemporary personality psychology, in general, and the density-distribution approach to personality, in particular. A central purpose of this project was to examine whether character traits can be meaningfully described as density distributions of momentary virtue states. The overall presumption was that because the same person manifests different degrees of virtuous behavior on different occasions, a person's virtuous behavior over time forms a distribution that is marked by a relatively stable mean and more or less variability around this mean (cf. Fleeson, 2001).

This project focused in particular on the variability and stability of behavior that is relevant to six broad virtue domains proposed in the Values in Action (VIA) classification of character strengths by Peterson and Seligman (2004)—namely, *wisdom, courage, humanity, justice, temperance,* and *transcendence*. Peterson and Seligman described these virtues as core characteristics valued by moral philosophers and religious thinkers across time and cultures. Character traits fall under these six broad virtue domains and were conceptualized as the psychological processes and mechanisms by which virtues may be achieved. For example, a person may display the virtue of *transcendence* via a variety of traits including gratitude, zest, hope, humor, spirituality, or appreciation of beauty. These character traits are supposed to be conceptually quite

similar to broad personality traits and have been defined as "positive traits reflected in thoughts, feelings, and behaviors. They exist in degrees and can be measured as individual differences" (Park, Peterson, & Seligman, 2004, p. 603; Noftle, Schnitker, & Robins, 2011).

Adopting Fleeson's density-distribution approach, the presumption was that nearly all levels of character traits reflecting the six VIA virtues were regularly expressed in an individual's daily behavior, implying a large degree of intra-individual variability in virtuous behavior. At the same time, the average level of and the variability in virtuous behaviors typically shown by an individual should be highly consistent and predictable. These aggregated indicators (i.e., the mean-level and the standard deviation of the distribution) were expected to qualify as relatively stable "trait-like" individual difference characteristics. In summary, the first goal of the project was to test whether it is useful for the six broad VIA virtues to be described as density distributions of virtue states as a whole.

On the condition that the density-distribution approach offers a suitable model to capture both the inconsistent and consistent parts of character traits, the project further aimed to identify potential antecedents of intra-individual variability in virtuous behavior. Specifically, on the basis of a view of individuals as actively reacting to role contexts and situational demands (Allport, 1937; Fleeson, 2007; Mischel & Shoda, 1998; Moskowitz, 1982; Wood & Roberts, 2006), it was proposed that individuals often change their virtue states as a function of the social role context they are currently acting in. A case in point might the manager who is expected to act in an assertive, tough-minded way on the job but is expected to act very differently in her role as a mother (Heller et al., 2007).

The Project: "Loving Mum and Tough Career Girl . . ."

To test the propositions just outlined empirically, an experience-sampling study was conducted in which a sample of 83 working parents described their current virtue states six times per day over a period of 10 days. Equipped with smartphones that were programmed with a mobile data collection program, the sample provided more than 4,300 momentary behavioral reports while while being in the role of the parent vs. being at work.

Each behavioral report included questions about the participants' momentary expression of virtue states that can be organized under the six broad virtue domains proposed in the VIA classification (Peterson & Seligman, 2004). These questions were the same format as standard adjective-based personality measures with the exception that rather than describing themselves in general, participants were asked to describe their behavior, thoughts, and feelings during the previous hour (for a similar approach, see Bleidorn, 2009; Bleidorn & Peters, 2011; Fleeson, 2001, 2007).

For example, to measure participants' state level of *wisdom*, they were asked to indicate on a bipolar rating scale which of the following adjective item-pairs was better suited to describe their behavior, thoughts, and feelings during the previous 60 minutes: (1) imaginative versus uninspired, (2) interested versus uninterested, (3) inquisitive versus indifferent, and (4) smart versus unwise. In addition to the momentary reports of virtue states, participants also reported on their momentary positive and negative affect, their role context (parent vs. job) as well as on the number of other persons with whom they had interacted during the previous 60 minutes (see Bleidorn & Denissen, 2015, for a detailed description of the study design and experience-sampling measures).

The assessment of individual's momentary expression of virtue states is important for at least two reasons. First, it addresses a basic question about the nature of character traits—namely, how virtues and character traits manifest in everyday behavior and whether individuals differ in such manifestations. Second, the direct assessment of character-relevant behavior over time and across different situations allows for a testing of the conception of character traits that is inclusive of both consistency and variability in character-relevant behavior. In particular, the degree to which inter-individual differences in aggregated levels of virtue states are stable and predictable informs the question of whether broad traits are useful to describe general levels of virtuous behavior. On the other hand, the amount of intra-individual variability in virtue states will indicate the degree to which situational accounts of character are useful to explain an individual's momentary expression of virtues in a particular context.

HOW MUCH INTRA-INDIVIDUAL VARIABILITY IS PRESENT IN CHARACTER-RELEVANT BEHAVIOR?

The first hypothesis derived from the proposed density-distribution model of character was that there is a sizeable amount of intra-individual variability in character-relevant behavior (i.e., in virtue states). Intra-individual variability describes the degree to which a person acts differently on different occasions. In contrast, inter-individual differences describe the degree to which people differ from each other. Taken together, both intra-individual variability and inter-individual differences add up to the total variability in virtue states (i.e., variability across individuals and across occasions). State-of-the-art statistical techniques, such as multilevel modeling (e.g., Hox, 2002), can be used to decompose this total amount of variability into the relative portions due to intra-individual variability (i.e., within persons) versus inter-individual variability (i.e., between persons). In other words, these models estimate the relative degree to which people differ from themselves from moment to moment as compared to the degree to which they differ from each other in a certain virtue domain.

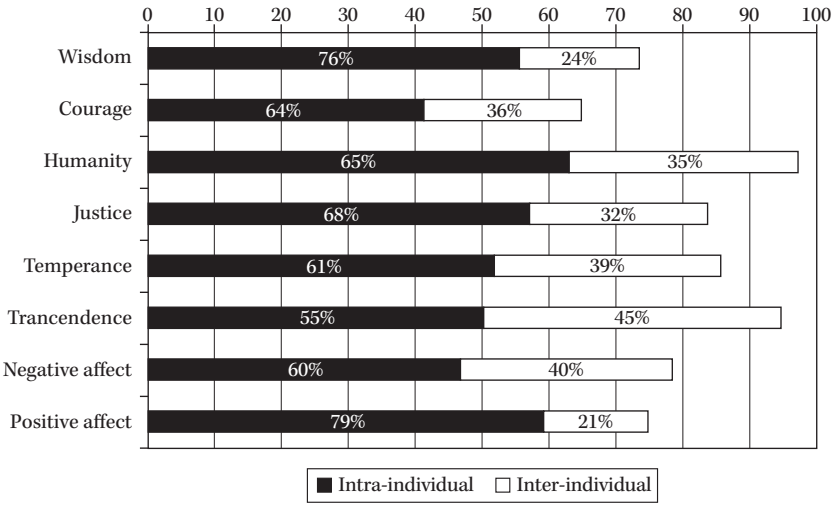

FIGURE 5.2 Absolute and relative amounts of intra-individual variability and inter-individual variability in virtue states and momentary affect.

As can be seen in figure 5.2, between 55% and 76% of the total variability in virtue states occurred within persons, whereas 24% to 45% occurred between persons; that is, individuals differed at least about the same, but generally even more from moment to moment in their virtue states than they differed from each other. In line with a density-distribution approach to character, the average individual showed nearly all levels of *wisdom, courage, humanity, justice, temperance*, and *transcendence* in the course of only a few days. To illustrate this, figure 5.3 shows the density distributions of the six virtue states for a random participant who completed a total of 57 experience-sampling reports over the course of 10 days.

Figure 5.2 also shows the absolute and relative amounts of intra-individual variability and inter-individual differences in positive affect and negative affect as a benchmark. Intra-individual variability in momentary affect is a useful comparison standard because affective experiences are so variable that they are primarily conceived as states (Fleeson, 2001; Bleidorn & Peters, 2010). As can be seen in figure 5.2, the amount of intra-individual variability in the six virtue states is actually as large as the intra-individual variability in positive and negative affect. That is, people seem to switch in their virtue states just as much as they switch from being in a good mood to being in a bad mood.

IS INTRA-INDIVIDUAL VARIABILITY IN CHARACTER-RELEVANT BEHAVIOR MEANINGFUL?

The sizeable amount of within-person variability in virtue states raises the question why individuals routinely and regularly shift in their virtuous

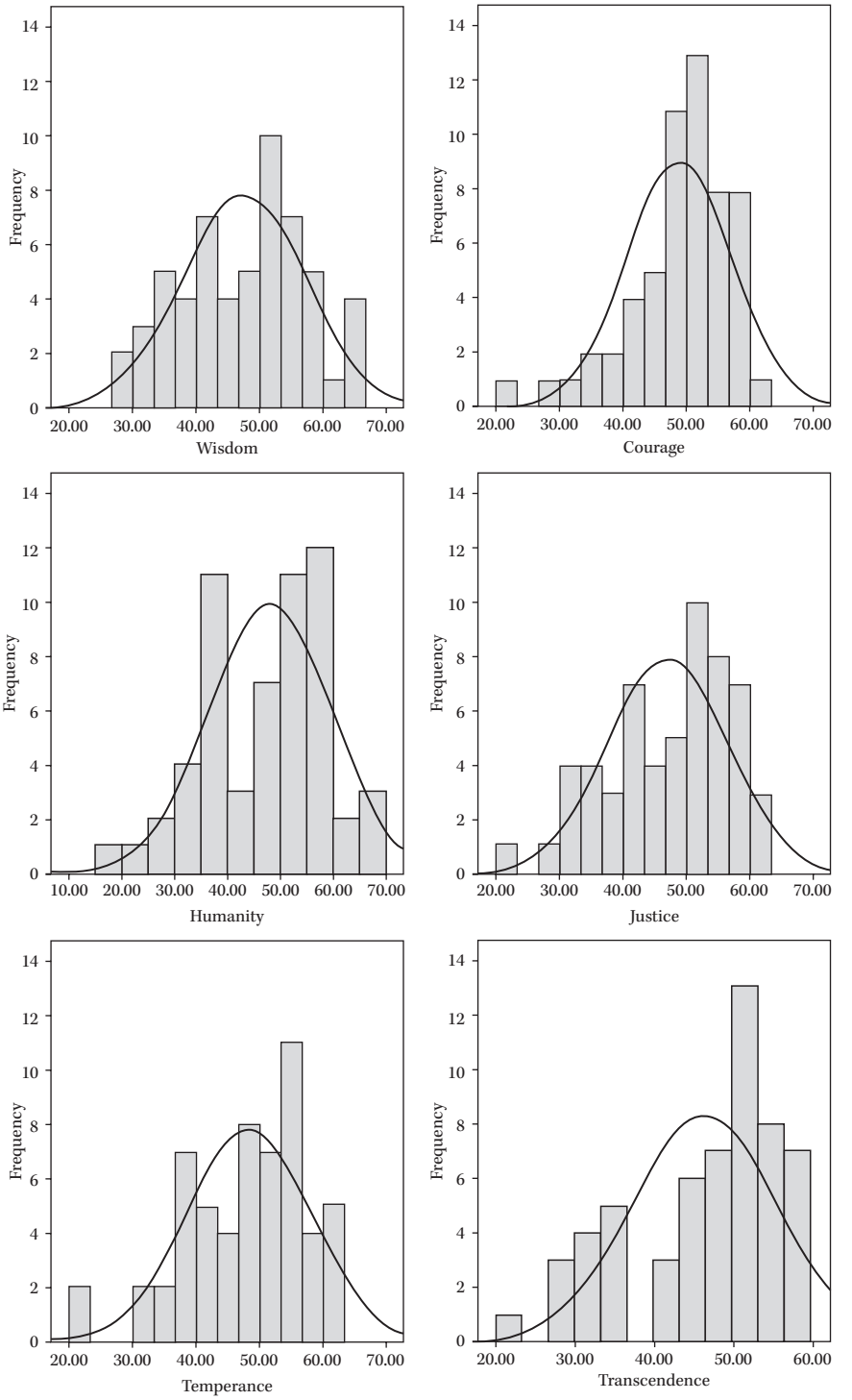

FIGURE 5.3 Density distributions of virtue states for a randomly selected participant who competed $N = 57$ experience-sampling reports over the course of ten days.

behavior. Does the large degree of intra-individual variability reflect a flexible and maybe even adaptive discrimination of situational demands or is it merely the result of random fluctuations and measurement error? This is an important question, because a further examination of intra-individual variability in virtue states would be worthwhile only if this variability is meaningful and predictable rather than random error or capricious variance (Fleeson, 2007).

From a contingent (or situational) perspective, it can be proposed that a large part of the observed variability in virtuous behavior is likely to reflect variability in situational cues. In other words, virtuous everyday behavior should at least partly vary as a function of changes in situational cues and contextual demands (Fleeson, 2001). The presumption in the present project was that working parents flexibly adapt their levels and sets of virtue states to their current social-role context. For example, working parents were expected to show higher levels of *humanity* (e.g., acting more warm-hearted, friendly, and grateful) while they were in the role of the parent as compared to being at work. And indeed, the experience-sampling data supported this prediction and suggested that working parents' variability in virtue states was not only random variance, but could also be partly explained by the social role context in which they were acting. In the parent role, the strongest links were found with virtue states that fall within the domains of *humanity* and *temperance*; in the job role, working parents most often showed virtue states that fall within the domains of *justice* and *wisdom* (for a detailed description of these results, see Bleidorn & Denissen, under submission).

At the same time, the results of the present project clearly suggested that individuals differ in their sensitivity to situational cues, such as role demands. For example, some working parents did not change or even decrease their momentary expression of *humanity* while in the parent role. These individual differences in sensitivity or reactivity to situational cues have been reported in previous experience-sampling studies on personality states (Bleidorn, 2009; Fleeson, 2007) and have been described as person x situation interactions (Magnusson & Endler, 1977), conditional if-then signatures (Shoda, Mischel, & Wright, 1994), or situation-based contingencies (Fleeson, 2007). The findings of the present study strongly support these *interactionist* positions by showing that how a person will act in a given situation is not only a function of the individual's character trait level or of the current situation but, rather, depends on how the person uniquely responds to the salient cues in a given situation (Fleeson, 2007); that is, a person's unique set of behavioral contingencies may be best considered as a part of his or her character that predicts how he or she may act on future occasions (e.g., Julia is usually assertive and dominant at work, but understanding and patient as a mother; Tom is usually fair as a father, but a fraud at work).

The question remains why individuals differ in these contingencies, and the next important step for future research would be to explain the origin

and development of these individual differences. Why do some people flexibly adapt their virtue states to the current role context, while others remain more or less stable? Answers to this question will not only inform interactionist approaches to personality but also shed further light on the processes underlying the manifestation of character traits and virtues in an individual's everyday behavior.

ARE INTER-INDIVIDUAL DIFFERENCES IN CHARACTER-RELEVANT BEHAVIOR CONSISTENT?

Reliable and consistent inter-individual differences in the expression of virtue states are important to the viability of the trait concept. Although single virtue states were highly variable, it was proposed that parameters that meaningfully describe the location and size of the entire distribution of virtue states would prove to be relatively stable individual difference characteristics. Referring to previous research on personality states, the presumption was that (1) the average tendency (i.e., the mean-level) and (2) the size (i.e., the standard deviation) of the virtue distributions would prove to be relatively stable "trait-like" variables (e.g., Bleidorn, 2009; Fleeson, 2001).

To test the first hypothesis, each participant's momentary experience-sampling reports were randomly split into two equal halves. For each half, the mean-level was calculated for each of the six virtue domains, indicating the average tendency of a person's virtuous behavior during that period of time. The mean-levels of each person's two halves were then compared with each other to test how stable the average levels of virtuous behavior were over the two time periods. The results of these comparisons suggested that the mean-levels of all six virtue states were highly stable over the two randomly assembled time periods. Figure 5.4 shows a scatter plot in which each participant's average level of virtue states in the domain of *wisdom* during one half of the study period is graphed as a function of that person's average *wisdom* levels during the other half of that period. This graph shows that different people had different mean-levels of state *wisdom*; but the two mean-levels from the same person were highly similar to each other. The scatter plots for the other five virtue domains looked highly similar to the one shown in figure 5.4, suggesting that people's average levels of virtuous behavior in all six VIA domains were highly stable across time (for details, see Bleidorn & Denissen, 2015); that is, in line with previous findings on personality states, people differed from each other in their average levels of virtue states, and these differences were precisely matched in the two independent halves of the data (cf. Fleeson, 2001).

The second hypothesis was that not only the average level but also the size of the distributions would prove to be a relatively stable individual difference characteristic. A common statistic to describe the size of frequency

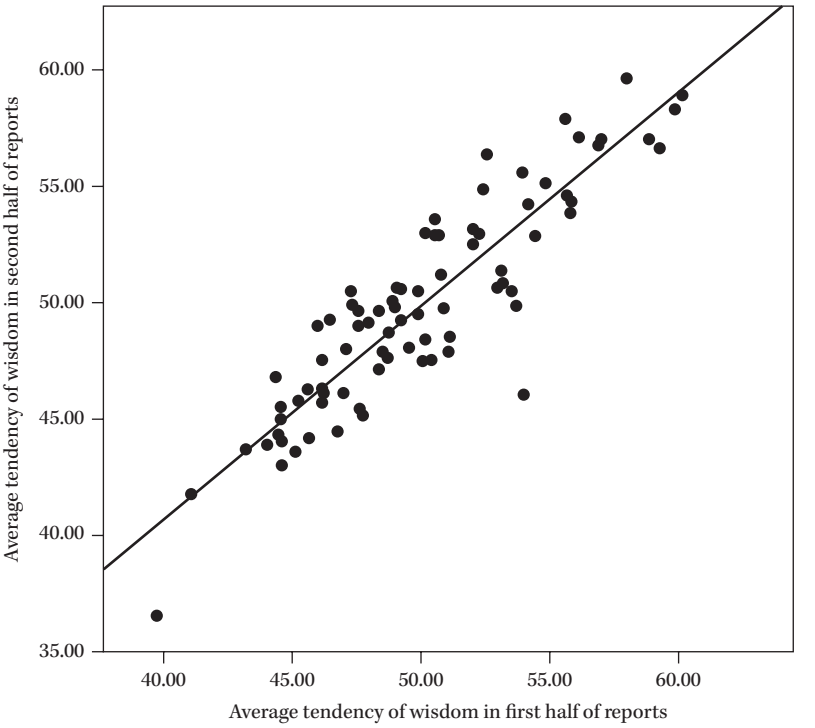

FIGURE 5.4 **Stability in virtuous behavior over time.** Each point in this graph represents one person's average tendency of state wisdom in two halves of the experience-sampling data (each half comprised ca. 25–30 different occasions). How much wisdom people displayed in their behavior during one half of the data-collection period was highly similar to how much wisdom they showed during the other half of the period.

distributions is the standard deviation. It is possible to calculate a standard deviation for each person describing the amount that his or her states differed across occasions. A large standard deviation indicates that the individual's state levels varied widely; whereas a low standard deviation indicates that the individual behaved very similarly on most occasions. To test the second hypothesis, each participant's momentary experience-sampling reports were again randomly split into two equal halves. For each half, the standard deviation was calculated for each of the six virtue domains, indicating the average range of a person's virtuous behavior during that period of time. The standard deviations of each person's two halves were then compared with each other. Similar to the average level, the size of the density distributions of virtue states also turned out to be highly stable across the two halves of the data. In other words, individuals differed from each other in how much they varied in their virtue states and these between-person differences in variability were highly stable across the two independent halves of experience-sampling reports (for details, see Bleidorn & Denissen, 2015).

In sum, both the average level and the size of the density distributions of virtue states were stable and highly predictable. This suggests that individual differences in the six VIA virtues can be meaningfully described as individual differences in density distributions of virtue states.

Conclusion

The project described in this chapter was dedicated to move the study of character beyond the person-situation debate by endorsing both the consistent and dynamic nature of character traits and virtuous behavior. To this end, how character traits and virtues manifest in people's everyday behavior was examined.

The findings of this project support the general proposal that character traits can be adequately described as density distributions of virtue states. That is, individuals indeed show a wide range of virtue states in their every-day behavior, suggesting a large degree of intra-individual variability (i.e., inconsistency) in virtuous behavior. This variability was not entirely random, but could be partly explained by the social role context the person was acting in. This suggests that the typical individual flexibly adapts his or her character-relevant behavior to salient situational cues, while there are still substantial inter-individual differences in how people typically react to the same situational cues. At the same time, people's average tendency to express certain virtue states as well as their variability in virtue states turned out to be highly consistent individual difference characteristics. These two parameters can be considered as trait-like variables that are useful to provide accurate descriptions of an individual's behavioral distribution.

Three conclusions can be derived from these results. First, the trait concept in character psychology is useful and deserves more attention in contemporary personality psychology. In fact, more research will be needed to examine the consequences of character traits and virtues for important life outcomes, such as marital satisfaction, distress, occupational success, or health-related outcomes (Noftle et al., 2011). However, scholars need to adopt an understanding of traits that captures the large variability in people's momentary behavior and acknowledge that character traits are best used for predicting behavior trends rather than singular behaviors (cf. Fleeson, 2004).

Second, because character traits cannot explain why a person acts differently on different occasions, scholars need to incorporate the dynamic (or inconsistent) aspects of virtuous behavior into models of character traits. Future research is needed to investigate the sources, conditions, and mechanisms of the dynamic changes and large variability in every-day virtuous behavior. What are the most relevant situational cues and individual states that may trigger increases or decreases in virtuous behavior?

A third conclusion is that other parameters of behavioral distributions may be important to character traits and our understanding of a person's character in general (Fleeson, 2001, 2004). For example, the present project suggested that individuals differ in how variable their virtuous behavior is over time as well as in their sensitivity to situational cues. Future research is needed to study the origins, development, and potential consequences of these individual differences in momentary expressions of character traits.

Finally, getting back to the initial question—there still is no definite answer to the question of how we should interpret the conduct of former US president Bill Clinton. From a density-distribution perspective on character traits, it could be argued that Clinton has shown a wide range of virtue state levels throughout his career, and that his average or typical behavioral tendency probably provides a more reliable description of his general character than do various single acts and episodes.[2]

References

Alfano, M. (2013). *Character as moral fiction*. New York: Cambridge University Press.

Allport, G. W. (1921). Personality and character. *Psychological Bulletin, 18*(9), 441–455.

Allport, G. W. (1927). Concepts of trait and personality. *Psychological Bulletin*, *24*(5), 284–293.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.

Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *European Journal of Personality*, *23*(6), 509–530.

Bleidorn, W., & Denissen, J. J. A. (2015). *Virtues in action – the new look of character traits. British Journal of Psychology*. doi: 10.1111/bjop.12117

Bleidorn, W., Kandler, C., Caspi, A. (2014). The behavioral genetics of personality development in adulthood: Classic, contemporary, and future trends. *European Journal of Personality, 28*(3), 244–255.

Bleidorn, W., Kandler, C., Riemann, R., Angleitner, A., & Spinath, F. M. (2009). Patterns and sources of adult personality development: Growth curve analyses of the NEO PI-R scales in a longitudinal twin study. *Journal of Personality and Social Psychology*, *97*(1), 142.

Bleidorn, W., & Peters, A. L. (2011). A multilevel multitrait-multimethod analysis of self- and peer-reported daily affective experiences. *European Journal of Personality*, *25*(5), 398–408.

Clinton, B. (2005). *My life*. New York: Random House.

[2] This project was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Thanks to Will Fleeson and Christian B. Miller for providing useful comments on an earlier version of this manuscript.

Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*(1), 100–108.

Donnellan, M. B., Lucas, R. E., & Fleeson, W. (2009). Introduction to personality and assessment at age 40: Reflections on the legacy of the person–situation debate and the future of person–situation integration. *Journal of Research in Personality*, *43*(2), 117–119.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.

Epstein, S., & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin, 98*(3), 513–537.

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027.

Fleeson, W. (2004). Moving personality beyond the person-situation debate the challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13*(2), 83–87.

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, *75*(4), 825–862.

Fleeson, W. (2012). Perspectives on the person. In Kay Deaux & Mark Snyder (Eds.), *The Oxford handbook of personality and social psychology* (pp. 33–63). New York: Oxford University Press.

Fleeson, W., & Noftle, E. (2008a). The end of the person–situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*(4), 1667–1684.

Fleeson, W., & Noftle, E. E. (2008b). Where does personality have its influence? A supermatrix of consistency concepts. *Journal of Personality*, *76*(6), 1355–1386.

Fleeson, W., & Wilt, J. (2010). The relevance of Big Five trait content in behavior to subjective authenticity: Do high levels of within-person behavioral variability undermine or enable authenticity achievement? *Journal of Personality*, *78*(4), 1353–1382.

Funder, D. C. (2008). Persons, situations and person-situation interactions. In O. P. John, R. Robins & L. Pervin (Eds.), *Handbook of personality* (3rd ed., pp. 568–580). New York: Guilford.

Goren, P. (2002). Character weakness, partisan bias, and presidential evaluation. *American Journal of Political Science*, *46*(3), 627–641.

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology & Penology, 1*, 69–97.

Hartshorne H, & May A. (1928). *Studies in the nature of character: Studies in deceit*. New York: Macmillan.

Hartshorne, H., & May, M. A. (1930). A summary of the work of the character education inquiry. *Religious Education, 25*(7), 607–619.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Quantitative methodology series. Mahwah, NJ: Erlbaum.

Heller, D., Komar, J., & Lee, W. B. (2007). The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin, 33*(6), 898–910.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research, 3*, 114–158.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43*(1), 23–34.

Kretschmer, E. (1925). *Physique and character.* Oxford, England: Harcourt, Brace.

Langer, G. (2001). Poll: Clinton legacy mixed. ABC News, January 17. Retrieved from http://abcnews.go.com/Politics/print?id=120952.

Magnusson, D. & Endler, N. S. (1977). *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, NJ: Erlbaum.

Moskowitz, D. S. (1982). Coherence and cross-situational generality in personality: A new analysis of old problems. *Journal of Personality and Social Psychology*, *43*(4), 754.

McCabe, K. O., & Fleeson, W. (2012). What is extraversion for? Integrating trait and motivational perspectives and identifying the purpose of extraversion. *Psychological Science*, *23*(12), 1498–1505.

McNiel, J. M., & Fleeson, W. (2006). The causal effects of extraversion on positive affect and neuroticism on negative affect: Manipulating state extraversion and state neuroticism in an experimental approach. *Journal of Research in Personality*, *40*(5), 529–550.

Milgram, S. (1974). *Obedience to authority: An experimental view*. London: Tavistock Publications.

Miller, C. B. (2013). *Moral Character: An Empirical Theory*. New York: Oxford University Press.

Miller, C. B. (2014). *Character and moral psychology*. New York: Oxford University Press.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology*, *49*(1), 229–258.

Newcomb, T. M. (1929). *The consistency of certain extrovert-introvert behavior patterns in 51 problem boys.* New York: Columbia University, Teachers College, Bur. Publ.

Noftle, E. E., Schnitker, S. A., & Robins, R. W. (2011). Character and personality: Connections between positive psychology and personality psychology. In In K. M. Sheldon, T. B. Kashdan, & M. F. Steger (Eds.), *Designing positive psychology: Taking stock and moving forward* (pp. 207–227). Oxford, England: Oxford University Press.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401–421.

Park, N., Peterson, C., & Seligman, M. E. (2004). Strengths of character and well-being. *Journal of Social and Clinical Psychology*, *23*(5), 603–619.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.

Roberts, B. W., & Caspi, A. (2001). Personality development and the person-situation debate: It's déjà vu all over again. *Psychological Inquiry*, *12*(2), 104–109.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4), 313–345.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. Philadelphia, PA: Temple University Press.

Schwesinger, G. C. (1926). Slang as an indication of character. *Journal of Applied Psychology, 10*(2), 245–253.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*(1), 5–34.

Shoda, Y., W. Mischel, and J. Wright. (1994). "Intraindividual Stability in the Organization and Patterning of Behavior: Incorporating Psychological Situations into the Idiographic Analysis of Personality." *Journal of Personality and Social Psychology* 67: 674-687.

Wood, D., & Roberts, B. W. (2006). Cross-Sectional and Longitudinal Tests of the Personality and Role Identity Structural Model (PRISM). *Journal of Personality*, *74*(3), 779–810.

6 }

Character Traits in the Workplace

A THREE-MONTH DIARY STUDY OF MORAL AND IMMORAL ORGANIZATIONAL BEHAVIORS

Taya R. Cohen and A. T. Panter

What is moral character? Can we measure it by asking people about their thoughts, emotions, and behaviors? Do people's self-perceptions of character correspond to other people's perceptions of them? Can measuring moral character with self-reports and/or observer reports help us predict and explain employees' harmful and helpful behaviors in the workplace? These are some of the questions we set out to study in the Work Experiences and Character Traits project (WECT Project; www.WECTProject.org). Our goal in designing this project was to investigate how moral character, personality, emotions, and treatment by managers and coworkers affect the frequency with which workers engage in ethical and unethical behavior at their jobs. The research is at the intersection of social/personality psychology and organizational behavior. As such, we think it has the potential to offer important theoretical and practical insights about character to both fields. Our hope is that the interdisciplinary nature of the project will create bridges between psychological, philosophical, and organizational research on morality, character, behavioral ethics, and workplace deviance.

Guided by four key questions related to moral character initially raised by the Character Project at Wake Forest University, we investigated the following: (1) Are there reliable and meaningful individual differences in dispositions to think, feel, and act in morally relevant ways? (2) Is there evidence, including consistency evidence, for the existence of character traits? (3) If character traits do stand up to empirical scrutiny, are they best conceptualized as general traits, such as those in the Big Five approach (openness, conscientiousness, extraversion, agreeableness, and neuroticism), or are they better understood as hundreds of highly localized traits? If they are general traits, does the Big Five approach capture all of them? (4) How accurate is the

understanding people typically have of their own character traits? And how accurate is the understanding people have of other people's character traits?

To address these questions, we drew on empirical findings from the WECT Project and other related studies we have conducted (Cohen, Wolf, Panter, & Insko, 2011; Cohen, Panter, & Turan, 2013a; Cohen, Panter, Turan, Morse, & Kim, 2013b; Cohen, Panter, Turan, Morse, & Kim, 2014; Wiltermuth & Cohen, 2014; for a review of this work, see Cohen & Morse, 2014). Because the WECT Project results are so foundational to our emerging insights about moral character in the workplace, we begin by describing the methods used in this research, and then move on to discussing how our findings relate to the four key questions outlined above.

The WECT Project

In the WECT Project, 1,514 US adults with full-time employment participated in a three-month online diary study about "weekly experiences in the workplace." We conducted the study in two separate administrations. Data collection for the first iteration began in September 2011, and data collection for the second iteration began in January 2012. Both iterations had the same procedures. In the initial survey, participants first provided detailed information about their jobs and completed a battery of measures assessing various dimensions of character and personality. To measure "moral character," we asked participants about 22 individual differences that have been linked to ethical and unethical choices in prior research. These variables were also measured again in the final survey, which largely paralleled the initial survey (a complete description of the individual difference variables we investigated is provided in Cohen et al., 2014). These variables were the six HEXACO personality factors (honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience); guilt proneness, shame proneness, and guilt and shame action tendencies (guilt-repair orientation, shame-withdrawal orientation); empathic concern and perspective taking; moral identity internalization and symbolization; consideration of future consequences and future self-continuity; moral idealism and moral relativism; cognitive moral development; narcissism; Machiavellianism; and self-control.

Following the initial survey, participants completed an online survey once a week for 12 weeks in which they reported the frequency with which they performed organizational citizenship behaviors (OCBs) and counterproductive work behaviors (CWBs). The OCB items assessed ethical organizational behaviors (e.g., helping coworkers with work and personal responsibilities, coming in early or staying late without pay); whereas the CWB items assessed unethical organizational behaviors (e.g., verbally or physically abusing coworkers, stealing from one's employer). Prior to conducting the WECT project we conducted a

pilot study of approximately 400 working adults to verify our assumption that people judge acts of OCB to be "moral" and acts of CWB to be "immoral" (see Appendix in Cohen et al., 2014, for a description of the pilot study about employees' moral judgments of OCB and CWB).

In addition to measuring self-reported OCB and CWB, the weekly surveys also asked participants about how they were treated by managers and coworkers (e.g., the frequency of interpersonal conflict, discrimination, ostracism, and abusive supervision) and the extent to which they had experienced various emotions during the previous week, both at work and outside work. Finally, the WECT Project also included observer reports of personality and behavior. We collected observer reports halfway through the project by asking coworkers of the participants to complete an online survey about the target employees' guilt proneness, shame proneness, and the six HEXACO factors, as well as the target employees' CWB and OCB during the previous month (see Cohen et al., 2013b, for more information about the observer reports and related findings). Collecting the observer reports and self-reports of the same individual differences and work behaviors allowed us to examine the correspondence between employees' perceptions of themselves and coworkers' perceptions of them with regard to both personality traits and ethical and unethical work behaviors.

Participants in the WECT Project were recruited from an online panel administered by a survey research firm specializing in web-based research. These individuals were a diverse group of American adults. They lived in all 50 US states and their ages ranged from 18 to 71 years ($M = 39.3$ years; $SD = 11.4$). Half of them were women. The sample contained White (75.2%), Black (9.2%), Hispanic (5.5%), Asian (3.6%), and other/multiracial (6.3%) participants. The jobs and organizations these individuals worked in were also diverse. All 23 occupational categories classified by the US Bureau of Labor Statistics were represented in the sample. Employees worked in private, for-profit companies (66.6%); government (14.7%); non-profit organizations (10.6%); and some were self-employed (8%). The median income in the sample was $44,000, but there was wide variability ($M = $52,962; $SD = $43,547).

Participants received compensation for their participation. Those in the first iteration of the study (WECT 2011) received $53, and those in the second iteration (WECT 2012) received $37. Individuals who missed surveys or terminated their participation early received partial compensation based on the number of surveys they had completed. We structured the payments so that the initial and final payments were larger than the interim payments for the weekly surveys. We did this to incentivize participants to complete the entire study. In WECT 2011, participants received $6 for the initial survey, $1 for each weekly survey, and $35 for the final survey; the coworkers received $20 for completing the coworker survey. In WECT 2012, participants received $5 for the initial survey, $1 for each weekly survey, and $5 for the final survey;

coworkers' received $15 for the final survey. We downwardly adjusted payment rates in the second iteration of the data collection because we learned from our experience with the first iteration that financial incentives did not need to be large to attract and retain participants. Indeed, it was surprising to us that payments to participants did not need to be substantial to incentivize them to complete the repeated surveys. Based on responses participants provided in open-ended comment boxes at the end of each survey, we learned that by and large, the participants we surveyed were at least somewhat intrinsically motivated to participate in this research. They enjoyed being able to talk about their work experiences with a team of researchers from prestigious universities. Through our surveys, we were giving voice to employees who might not typically encounter such opportunities in their daily lives. We speculate that this may have been particularly important for those who had difficult work environments, such as an abusive boss or difficult coworkers—they were happy to have someone to talk about these issues with, even if only in an online survey format.

Having provided details about the methods we used in this research, we now turn our attention to addressing what we learned from the project about moral character and ethical work behaviors.

Question 1: Are there reliable and meaningful individual differences in dispositions to think, feel, and act in morally relevant ways?

We included multiple potential measures of character in our study because our key goal was to investigate the importance of different traits in determining a person's likelihood of behaving ethically. This is in contrast to most extant research on character, which tends to focus on a single character trait or a small set of character traits. We conducted a variety of different statistical analyses to analyze the measures, including correlation analysis, regression analysis, latent profile analysis, and factor analysis (see Cohen et al., 2014, for a complete description of these analyses and results). Based on these analyses, our response to Question 1 is an unequivocal yes—there are indeed reliable and meaningful individual differences in dispositions to think, feel, and act in morally relevant ways. Of particular importance, most of the individual differences measured in the initial survey predicted subsequent ethical and unethical behavior in the workplace (i.e., OCB and CWB), supporting the claim that individual differences in moral character exist.

Specifically, our data suggest that individuals with high levels of moral character can be characterized in a certain profile. This profile can be described with a tripartite model (Cohen & Morse, 2014). According to this model, moral character consists of a variety of virtuous traits, some of which influence

the motivation to be moral (i.e., one's desire to do right and to avoid doing wrong); others of which influence the ability to be moral (i.e., one's capacity to do right and to avoid doing wrong); and others that influence moral identity (i.e., one's identity as a good vs. bad person). The three components of our tripartite model are not orthogonal, so it is certainly possible, and likely, that some character traits relate to more than one component.

In our view, the motivational component of moral character encompasses traits related to the consideration of the needs and interests of others as well as of how one's own actions affect others. Examples of traits that we think relate to the consideration of others include honesty-humility, low Machiavellianism, empathic concern, perspective taking, guilt proneness, and guilt-repair orientation.

Honesty-humility is a broad personality dimension indicative of a person's fairness, sincerity, modesty, and greed-avoidance (Ashton & Lee, 2007, 2009). A person who is high on honesty-humility is fair and genuine in dealing with others, even when doing so could put him or her at risk for being taken advantage of. For example, this person would strongly agree with the following statements: "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed"; and "Having a lot of money is not especially important to me."

Machiavellianism refers to the extent to which a person is manipulative and dishonest in their interactions with others (Christie & Geis, 1970). It is the opposite of honesty-humility in many ways. A person who is high on Machiavellianism treats other people as objects and exploits them for personal gain. A person who is low on Machiavellianism values honesty and believes that people get ahead in the world by leading moral lives. For example, a person who is low on Machiavellianism would strongly disagree with the statement: "Never tell anyone the real reason you did something unless it is to your benefit to do so."

Empathic concern refers to the affective component of empathy; whereas perspective taking refers to the cognitive component of empathy (Davis, 1983). Both are relevant to moral character. A person who is high on empathic concern experiences feelings of sympathy, compassion, and concern for others' well-being. For example, this person would strongly agree with the following statement: "I often have tender, concerned feelings for people less fortunate than me." A person who is high on perspective taking adopts the perspectives of other people and tries to understand how they think and feel. Such persons imagine themselves in "other people's shoes." For example, they would strongly agree with the following statement: "I try to look at everybody's side of a disagreement before I make a decision."

Guilt proneness can be thought of as the strength of one's conscience—would the person feel badly about committing a transgression, such as harming another person, even if no one knew about what they did (Cohen et al., 2011)?

A person who is high on guilt proneness feels bad about wrongdoing, even when the wrongdoing is private. For example, this person would respond "very likely" to the following question: "After realizing you have received too much change at a store, you decide to keep it because the sales clerk doesn't notice. What is the likelihood that you would feel uncomfortable about keeping the money?"

Guilt-repair orientation is related to guilt proneness (Cohen et al., 2011). A person who is high on this trait engages in corrective behaviors following wrongdoing, even when the wrongdoing was private. That is, they anticipate that they would make reparations for past mistakes by changing their future behavior, even if no one knew about their mistakes. For example, they would respond "very likely" to the following question: "You reveal a friend's secret, though your friend never finds out. What is the likelihood that your failure to keep the secret would lead you to exert extra effort to keep secrets in the future?"

The ability component of the model comprises various traits related to the regulation of one's behavior, specifically with reference to regulating behavior that may have positive short-term consequences but negative long-term consequences for one's self or others. Personality characteristics relevant to this aspect of the model are indicative of an ability to act ethically and to refrain from acting unethically should the person be motivated to act in such a way. If motivations or intentions to behave in an ethical manner are absent, it is unlikely that self-regulatory abilities, on their own, would facilitate ethical behavior. However, when motivations and intentions to be moral are present, individual differences in self-regulation can foster helpful behaviors and deter harmful behaviors. Examples of traits that relate to self-regulation include conscientiousness, self-control, and consideration of future consequences.

Conscientiousness is a broad personality dimension indicative of a person's organization, diligence, perfectionism, and prudence (Ashton & Lee, 2007, 2009). Persons who are high on conscientiousness organize their time and their physical surroundings, work in a disciplined way toward their goals, strive for accuracy and perfection in their tasks, and deliberate carefully when making decisions. For example, they would strongly agree with the following statements: "I plan ahead and organize things, to avoid scrambling at the last minute"; and "I always try to be accurate in my work, even at the expense of time." Conscientiousness is the broad personality factor linked most closely to the ability to self-regulate—an ability that is usually required for forgoing the temptation of short-term gains that generally result from unethical behavior. Self-control is related to conscientiousness, and refers to an ability to override desires and work effectively toward long-term goals (Tangney, Baumeister, & Boone, 2004). A person who is high on self-control refrains from acting on undesired behaviors and adapts oneself to act in a desirable way. This person

would strongly agree with the following statement: "I am good at resisting temptation."

Consideration of future consequences refers to the extent to which individuals think about the long-term consequences of their behavior and modify their future behavior accordingly (Strathman et al., 1994). A person who is high on consideration of future consequences would strongly agree with the following statement: "I consider how things might be in the future, and try to influence those things with my day to day behavior." Consideration of future consequences facilitates self-control and conscientiousness, and is important for self-regulation.

The third category in the model is moral identity. A person with a highly internalized moral identity has a deep concern about being a moral person and viewing himself or herself accordingly (Aquino & Reed, 2002). Specifically, a person who scores highly on measures of moral identity internalization would feel very good to be a person who possesses the following characteristics: "caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, and kind." This person would strongly agree that being someone who has these characteristics is an important part of who they are (for more on the development of this measure and why this list of moral identity characteristics were chosen, see Aquino & Reed, 2002). Moral identity is related to the motivation and ability elements of moral character in that a person could identify as a person who is very considerate of others (e.g., generous) and/or as a person who has a strong ability to self-regulate (e.g., hardworking). That is, they could identify with different aspects of moral character, some of which might be more relevant to the motivation to be moral and others of which might be more relevant to the ability to be moral.

Drawing on this model and our empirical results from the WECT Project, we can describe people with high levels of moral character as follows. They refrain from manipulating others (low levels of *Machiavellianism*; Christie & Geis, 1970). They consider other people's perspectives and feelings (high levels of *perspective taking* and *empathic concern*; Davis, 1983). They value integrity and want to see themselves as moral (high levels of *internalized moral identity*; Aquino & Reed, 2002). They anticipate feeling guilty and engaging in repair-oriented actions when they do something wrong (high levels of *guilt proneness* and *guilt-repair orientation*; Cohen et al., 2011). They are disciplined, prudent, organized, and good at resisting temptations (high levels of *conscientiousness* and *self-control*; Ashton & Lee, 2009; Tangney, Baumeister, & Boone, 2004). They think about future consequences of their behavior and act accordingly (high levels of *consideration of future consequences*; Strathman, Gleicher, Boninger, & Edwards, 1994). And in general, they are characterized as being sincere, modest, and fair (high levels of *honesty-humility*; Ashton & Lee, 2009). All the traits just described were found to predict subsequent

counterproductive and citizenship behaviors, as reported by the employees themselves and by their coworkers (Cohen et al., 2014).

Just as interesting as knowing the traits that predict ethical behavior is knowing the traits that do not. Not every potential variable we explored correlated with the character traits we identified above, nor did every variable predict CWB and OCB. We found two results particularly surprising—variables that seem to be relatively undiagnostic of moral character according to our analyses of how the variables we measured relate to one another and to ethical behavior. Identifying such variables helps us determine the scope of our model of moral character and provides insight into its structure.

Emotionality and cognitive moral development did not differentiate individuals with a moral profile versus those with immoral or amoral profiles. Emotionality is a broad personality dimension indicative of a person's fearfulness, anxiety, dependence, and sentimentality (Ashton & Lee, 2007, 2009). Prior research on five-factor personality models has suggested that emotional stability (a variable closely related to emotionality) predicts unethical behavior (Berry, Ones, & Sackett, 2007; Henle & Gross, 2013), but our study did not find any evidence of this. Indeed, of all the 22 variables we investigated, emotionality had the weakest relationships with the other moral character traits and with ethical and unethical work behaviors. It was uncorrelated with CWB and OCB and did not differentiate the employees in the sample in the same way as the variables we found to be important for understanding the profile of moral individuals.

Cognitive moral development also did not predict ethical and unethical behavior in our study. This construct is sometimes referred to as *moral reasoning* (Kohlberg, 1969; Rest, 1986) and is indicative of the complexity of a person's thoughts about difficult moral dilemmas. There are decades of social science research that focuses on moral reasoning ability, with the assumption that it relates to moral behavior (Kish-Gephart, Harrison, & Treviño, 2010; Narvaez, 2010). We found it particularly interesting that this variable did not predict helpful and harmful work behaviors and did not differentiate the employees in the sample in the same way as the variables that we found to be important for understanding the profile of moral individuals. Our view is that cognitive moral development likely predicts complex ethical decision-making, but is not particularly well-suited to predicting more common and mundane choices, such as the decision to commit CWB or OCB, where there is widespread agreement about the ethicality of the behaviors in question (Cohen et al., 2014).

Certainly, there are limitations of our study of moral character in the workplace and opportunities for future research that extends and clarifies our tripartite framework and related findings. We do not know whether the three components we have identified are comprehensive of the character landscape, nor do we know how these components fit together to determine moral behavior—is any one necessary or sufficient? Is their sum greater

than their parts? We hope that the framework we propose paves the way to further research by scholars in different fields that tests, extends, and challenges our current thinking on what moral character is and what moral character does.

Furthermore, we recognize that the conclusions we have drawn about the relationship between moral character and ethical behavior are based on survey data rather than experimentation. Thus, despite the causality implied by our assumptions and observations of how character traits influence work behaviors over time, our results are still subject to alternative causal explanations. Indeed, we believe the relationship between personality and behavior is a dynamic one and that causality may run in both directions. While being disposed to act in a particular way predicts relevant future behaviors, behaving in a particular way over time changes one's disposition. Currently, there is little research examining character and behavior dynamically over time, so there is much opportunity for future research that examines the causal structure of these relationships in more depth.

Question 2: Is there evidence, including consistency evidence, for the existence of character traits?

As with Question 1, based on our research in the WECT Project, we think the answer to this question is also a resounding yes. When we asked our participants to self-report about the moral character traits we described in the preceding section, these traits were all reasonably stable across the 14 weeks between the initial and final surveys in our project. Specifically, test-retest correlations (i.e., the correlation of a variable with itself at a later time point) were in the range of .60 to .70, similar in magnitude to test-retest correlations for other dimensions of personality that are less relevant to moral character, such as extraversion and openness to experience. Furthermore, self-reports of moral character traits predicted coworker reports of CWB and OCB five weeks later, and self-reports of CWB and OCB 12 weeks later (see Cohen et al., 2013b, 2014, for more on these results). Thus, our empirical findings indicate that self-reports of moral character have a strong degree of consistency in adults over a period of several months.

Despite the strong consistency over time that we observed in our data, we do not think that moral character is fixed. Like others who take a trait-theory view of moral character, we assume character traits are "stable and general, but also shaped by the individual's setting and thus capable of change" (Peterson & Seligman, 2004, p. 10). Indeed, we found differences in moral character, depending on the age of the respondent. Individuals who classified as low in moral character were, on average, seven years younger than those classified as high in moral character. We think this age difference makes sense in light of

our tripartite framework. As people age, they likely become more considerate of others and more skilled at self-regulation in a variety of settings; both of these developmental changes could make them more moral.

Question 3: If character traits do stand up to empirical scrutiny, are they best conceptualized as general traits, such as those in the Big Five approach (openness, conscientiousness, extraversion, agreeableness, and neuroticism), or are they better understood as hundreds of highly localized traits? If they are general traits, does the Big Five approach capture all of them?

Moral character, according to our data, is captured by both broad and narrow traits, many of which were described earlier. The two broad traits that we identified as particularly important elements of moral character were honesty-humility and conscientiousness. While conscientiousness is a factor in the Big Five, honesty-humility is not. Thus, the Big Five model does not capture all relevant personality and character traits. Instead, our results suggest that the HEXACO framework is superior to the Big Five framework when it comes to conceptualizing moral character (Ashton & Lee, 2007).

Research on the HEXACO model of personality indicates that there are six broad traits that encompass the entire landscape of personality. Honesty-humility is not fully captured by five-factor models. The six broad traits described by the HEXACO model can be detected in most, and possibly all, cultures throughout the world. More than a decade of research by Michael Ashton and Kibeom Lee supports this claim (for reviews, see Ashton & Lee, 2008; Lee & Ashton, 2012). In our work, honesty-humility predicted the frequency with which employees reported CWB and OCB in the workplace, as did conscientiousness (Cohen et al., 2013b, 2014), replicating previous work with the HEXACO scales and work behaviors (Bourdage, Lee, Lee, & Shin, 2012; Marcus, Lee, & Ashton, 2007). Honesty-humility and conscientiousness are also related to other narrower facets of moral character, such as guilt proneness and moral identity, which as described earlier, also predict ethical and unethical behaviors (Cohen et al., 2014).

In light of our results, we disagree with claims that character traits are highly localized to particularly settings. People with a tendency to behave unethically at work also have a tendency to behave unethically at home, and in a variety of other social situations (e.g., Cohen et al., 2011, 2013b, 2014). This does not mean that a person low in moral character will behave unethically in every aspect of his or her life. Rather, it means that we can measure and observe general tendencies in people's patterns of behaviors, and these patterns are often apparent regardless of the particular context in which you observe them.

Question 4: How accurate is the understanding people typically have of their own character traits? And how accurate is the understanding people have of other people's character traits?

Evidence related to this set of questions can be gleaned from the coworker reports that we were able to collect from a subset of the respondents in our weekly diary study. We asked employees to nominate a well-acquainted coworker who could complete self-ratings as well as ratings about the participant on several key moral character traits, including honesty-humility, conscientiousness, guilt proneness, and guilt-repair orientation. Because of time constraints, we could not administer the full battery of measures that we had the employees complete, so we focused on a subset with strong empirical and theoretical support for their relevance to moral character and behavior.

"Self-other agreement" refers to the extent to which an individual's self-ratings correspond to an observer's ratings of him or her. In our study, this agreement was captured by the correlation between a participant's self-report and his or her coworker's observer report. All the self-other agreement correlations that we examined were substantial (ranging from $r = .50$ to .60), with the exception of guilt-repair orientation, which was lower for reasons that are unclear ($r = .32$). These correlations were similar to those observed for other personality dimensions that we investigated (i.e., the other HEXACO traits and shame proneness), and the relationships held even after controlling for the similarity and assumed similarity between the participants and the coworkers (Cohen et al., 2013b). "Similarity" refers to whether two individuals rate themselves similarly. *Assumed similarity* refers to whether a perceiver sees himself or herself akin to how they see others; it captures projection of one's own personality onto others. The fact that we observed moderately strong self-other agreement correlations after controlling for similarity and assumed similarity correlations implies that neither projection of one's own traits on others (i.e., assumed similarity) nor actual similarity fully explains the consensus between self-ratings and observer ratings of moral character.

This consensus between how employees rate themselves and how well-acquainted coworkers judge them gives us confidence in the accuracy of both self-ratings and observer ratings. While neither method is free from limitations, the convergence between the two reporting methods goes a long way toward alleviating concerns about the problems associated with each method. Moreover, regardless of the reporting method, the moral character traits correlated negatively with unethical behavior. That is, coworker ratings of honesty-humility, conscientiousness, guilt proneness, and guilt-repair orientation all inversely predicted self-reports of CWB. Likewise, self-ratings of honesty-humility, conscientiousness, guilt

proneness, and guilt-repair orientation all inversely predicted coworker reports of CWB (Cohen et al., 2013b).

Conclusion

When we planned the WECT Project, we prioritized certain objectives. We wanted to build on our prior research in the areas of moral character, personality, emotions, and behavioral ethics, while recognizing that there were several unknown aspects of how moral character traits manifested themselves in real-world settings, such as one's work environment. We did not want to be limited by a single reporting source (usually self), a single operationalization of moral character, a single point in time, a single generalized context that ignored how most of the US adult population spend their time, and a single disciplinary perspective. We were motivated to bridge existing psychological research on morality with organizational research on behavioral ethics and workplace deviance.

By designing a research project that was consistent with those objectives, we were able to gain insight into the structure of moral character, including dimensions of personality and the self that are not as relevant to moral character. Moreover, we were able to show convergence in perceptions of moral character across multiple reporters, and demonstrate a strong relevance of moral character to understanding and predicting behaviors that occur in the workplace. Our current research builds on this work by focusing on methods that allow us to observe specific, relevant behaviors in the workplace. For example, we are looking at the extent to which trained observers can identify moral character traits and relevant behaviors in responses to job interview questions. We are investigating the types of interview questions that could elicit material that includes moral character information and testing the extent to which trained interviewers can reliably detect this information. Related to our tripartite model, we are considering an interviewee's motivation to be moral, their ability to provide strong examples of past moral behavior as perceived by others, and their own self-perception as having a moral identity. By establishing strong ties to observed behavior in workplace contexts, we can move forward in describing and testing the role that individual differences in moral character play in our social and work lives.[1]

[1]This work was made possible through the support of Grant 15519 from the Character Project at Wake Forest University and the John Templeton Foundation to Taya R. Cohen and A. T. Panter. The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the sponsors. We thank the members of the Character Project at Wake Forest University and attendees of the Character Project conferences for helpful feedback on this work. We also thank Nazli Turan, Lily Morse, and Yeonjeong Kim for valuable assistance with this research.

References

Aquino, Karl, & Reed, Americus. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. doi: 10.1037/0022-3514.83.6.1423

Ashton, Michael C., & Lee, Kibeom. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166. doi: 10.1177/1088868306294907.

Ashton, Michael C., & Lee, Kibeom. (2008). The HEXACO model of personality structure and the importance of the H Factor. *Social and Personality Psychology Compass, 2*, 1952–1962. doi:10.1111/j.1751-9004.2008.00134.x.

Ashton, Michael C., & Lee, Kibeom. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340–345. doi:10.1080/00223890902935878.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology, 92*(2), 410–424. doi: 10.1037/0021-9010.92.2.410

Bourdage, Joshua S., Lee, Kibeom, Lee, Jong-Hyun, & Shin, Kang-Hyun. (2012). Motives for organizational citizenship behavior: Personality correlates and coworker ratings of OCB. *Human Performance, 25*(3), 179–200. doi:10.1080/08959285.2012.683904.

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.

Cohen, T. R., & Morse, L. (2014). Moral character: What it is and what it does. In A. P. Brief & B. M. Staw (Eds.), *Research in Organizational Behavior*. Elsevier. doi:10.1016/j.riob.2014.08.003.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L. A., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, *107*(5), 943–963. doi: 10.1037/a0037245.

Cohen, T. R., Panter, A. T., & Turan, N. (2013a). Predicting counterproductive work behavior from guilt proneness. *Journal of Business Ethics*, *114*, 45–53. doi:10.1007/s10551-012-1326-2.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L. A., & Kim, Y. (2013b). Agreement and similarity in self-other perceptions of moral character. *Journal of Research in Personality*, *47*, 816—830. doi:10.1016/j.jrp.2013.08.009.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*, 947–966. doi:10.1037/a0022641.

Davis, Mark H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. doi: 10.1037/0022-3514.44.1.113

Henle, C. A., & Gross, M. A. (2013). Born to be deviant? An examination of the relationship between workplace deviance and employee personality. In S. M. Elias (Ed.), *Deviant and criminal behavior in the workplace* (pp. 50–76). New York: New York University Press.

Kish-Gephart, J. J., Harrison, D. A., & Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology, 95*(1), 1–31. doi: 10.1037/a0017103.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago, IL: Rand McNally.

Lee, Kibeom, & Ashton, Michael C. (2012). *The H factor of personality: Why some people are manipulative, self-entitled, materialistic, and exploitive—and why it matters for everyone*. Waterloo, Canada: Wilfrid Laurier University Press.

Marcus, B., Lee, Kibeom, & Ashton, Michael C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five, or one in addition? *Personnel Psychology, 60*(1), 1–34. doi: 10.1002/per.597

Narvaez, D. (2010). Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science, 5*(2), 163–181. doi:10.1177/1745691610362351.

Peterson, Christopher, & Seligman, Martin E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.

Rest, James R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*, 742–752. doi: 10.1037/0022-3514.66.4.742

Tangney, June Price, Baumeister, Roy F., & Boone, Angie Luzio. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–324. doi:10.1111/j.0022-3506.2004.00263.x.

Wiltermuth, S. S., & Cohen, T. R. (2014). "I'd only let you down": Guilt proneness and the avoidance of harmful interdependence. *Journal of Personality and Social Psychology*. doi:10.1037/a0037523.

7 }

The Mixed Trait Model of Character Traits and the Moral Domains of Resource Distribution and Theft

Christian B. Miller

Do most people today have the traditional moral virtues like honesty and courage? Or are most of us vicious people, possessing vices like dishonesty and cowardice?

One way to try to answer these questions is to look at the best experimental work being done in psychology. Over the past few years, I have spent a lot of time doing just this. And my conclusion is not one that I expected to arrive at, as I have come to believe that most people possess none of the traditional moral virtues and none of the traditional moral vices. Instead, they possess rather different character traits, which I called Mixed Traits.

In two recent books, I have articulated and supported my Mixed Trait model in some detail.[1] In doing so, I relied primarily on empirical work pertaining specifically to the moral domains of helping, physical aggression, lying, and cheating.

But naturally there are plenty of other moral domains besides these four, and so in this paper the goal is to extend my earlier discussion to two additional areas—resource distribution and theft. In doing so, I will consider whether the existing research is compatible with my Mixed Trait model, or whether instead it gives me reason to be concerned with how broadly applicable the model really is. My conclusion will be that the results are, so to speak, a mixed bag. With respect to distributing resources, some carefully conceived studies have been done which do in fact provide additional support for my model. But with

[1] Miller 2013, 2014.

respect to stealing, it turns out that we are badly in need of new behavioral studies before we can draw any reasonable conclusion about people's character.

Section 1 of this chapter offers a quick overview of my Mixed Trait model, saving the details for the two books. The remaining two sections consider the moral domains of resource distribution and theft.

1. A Brief Overview of the Mixed Trait Model

My Mixed Trait model is intended to offer a descriptive picture of what moral character looks like for most people today, at least in the West, where the relevant empirical research in psychology has primarily been done. Note that the model purports to be a *descriptive* picture about how we actually are, and is not in any way a normative account of what character traits we should strive to acquire. Furthermore, it is focused just on *moral* character traits and does not make any claims about, say, our epistemic, prudential, or artistic characters. Finally, it is a model pertaining to the here and now. It may also hold for earlier time periods, but since it's support only comes from psychological research done in the past fifty years, it would be too hasty to make such an inference.

Before spelling out the model in more detail, let me introduce a key assumption that does a lot of work in my thinking about character traits. Fortunately, the assumption is fairly uncontroversial among ethicists—namely, that for any given character trait, there are various normative standards it has to satisfy in order to qualify as a particular virtue. For instance, a character trait having to do with telling lies and cheating does not automatically qualify as the virtue of honesty. Obviously enough, it could be the vice of dishonesty instead. For the trait to count as a virtue, it has to fit the "job description" for that virtue. Here is an example of one of the many such criteria in the case of honesty:

> (1) A person who is honest will regularly refrain from cheating in situations where he is a free and willing participant and the relevant rules are fair and appropriate, even if by cheating he is assured of acquiring some benefit for himself.[2]

The same holds true for the vices—a character trait has to satisfy the relevant normative standards to qualify as a particular vice instead. Here is an example in the case of dishonesty:

> (2) A dishonest person does not have moral beliefs to the effect that cheating is wrong in general, as well as wrong in most particular instances of what are widely considered to be acts of cheating. Or, if he does

[2] See, e.g., Hursthouse 1999: 10; and Adams 2006: 121. Thanks to Tim Mawson for suggesting a revision to an earlier formulation.

> happen to have such beliefs, he will not care much about them, and they will not play a significant motivational role in his psychology.

Many other criteria could be specified as well, although I doubt we could ever come up with an exhaustive and precise list of all of them.

What enables a character trait to satisfy these standards for being a virtue or a vice? Briefly, the answer is that the beliefs, desires, emotions, and the like which make up that character trait have to satisfy the appropriate normative criteria themselves. If Jones has the virtue of honesty, for instance, then that is because (in part) he has certain correct *beliefs* about when to tell the truth. And he has appropriate *desires* to tell the truth in those instances. And he can reliably *recognize* when he is in a situation in which it is important to tell the truth, and he reasonably *weighs* different ways of telling the truth. At the same time, he is *not* strongly disposed to want to lie whenever it would be to his advantage and he could get away with it, since from any plausible ethical perspective that would not be a disposition which is an appropriate constituent of the virtue of honesty.

So one cannot count as having a virtue like honesty without first having, to a certain degree, the specific dispositions to form beliefs and desires relevant to honesty, and also not having (at least to a certain degree) the specific dispositions to form beliefs and desires relevant to dishonesty. A character trait qualifies as a particular virtue, then, provided that its underlying mental state dispositions are of that particular virtuous kind themselves.[3] Once a trait *does* satisfy the normative standards for being a particular virtue or vice, then it is not as if it is always possessed to the same extent. Two people can have the virtue of honesty, but each might have it to a different degree.[4] One might be only weakly honest as a person, while the other is deeply honest. To use language familiar from psychology, the virtues (and vices) are categorical traits—you have to meet certain standards to qualify as possessing them—but once you do, they are continuous traits which can show up anywhere along a wide continuum of more and less. So, in light of this discussion, there are two distinct questions that one can ask about a person's possession of the virtues (and vices): does she have a particular virtue, such as honesty, *at all*, and if she does, *to what degree* does she possess this trait?

To summarize, for any given moral domain where there is a virtue and a vice, we might say that there is a normative continuum on which a given person might fall. Figure 7.1 provides an illustration in the case of honesty and dishonesty. Where a person's character *actually is* on this continuum is an

[3] This is intended as a necessary condition; I remain neutral here as to whether it is sufficient for a character trait's qualifying as a particular virtue.

[4] The degree to which a person has a character trait is presumably some complex function of the degree to which he has the underlying mental state dispositions and the relations between them.

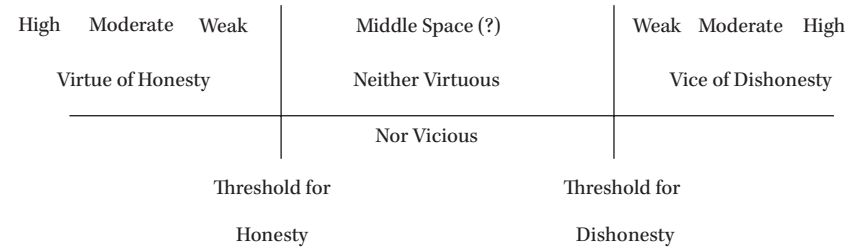

FIGURE 7.1 A Continuum for Honesty and Dishonesty.

empirical matter, but that there is such a continuum for a virtue like honesty is a conceptual and normative matter.

With this background in mind, let me briefly highlight a few features of my Mixed Trait model, again noting that the details are presented elsewhere. The negative side of my view can be stated rather briefly:

(3) On the basis of the available psychological research, there is good reason to think that most people do not have the appropriate mental state dispositions which enable them to qualify as possessing either any of the moral virtues or any of the vices to any degree, although a few people may so qualify.

This is a claim about what people are *not* like. According to this claim, published studies in psychology provide strong support for thinking that most people's mental lives do not meet the various normative standards to pass over the threshold for having virtues like honesty. But at the same time, those studies also suggest the same thing about the vices as well. So we should come to accept that, given the available evidence, most people today do not have the virtues and the vices.[5]

What can be said more positively about what most people *are* like? Here is where the Mixed Trait framework comes into play. For I claim it is also true that:

(4) Most people actually do possess traits of character pertaining to the different moral domains, and these traits consist of various interrelated mental state dispositions pertaining to that domain.[6]

How can this be? If these traits of character are not the traditional virtues or vices, then what are they like?

To see, let me use the example of cheating. In light of a series of recent studies, we can begin to infer what some of the mental state dispositions

[5] For the studies in question, see Miller 2013, 2014. Other philosophers have arrived at a similar conclusion, albeit using different studies than I tend to cite. See, for instance, Doris 1998, 2002; Harman 1999, 2000; Vranas 2005; and Alfano 2013.

[6] As already noted, this needs to be qualified to most people in the West in recent years. My picture of character *may* apply more broadly than this, but clearly a lot more research would need to be done first before I would feel comfortable making such a claim.

which seem to be widely held and which play a role in bringing about cheating behavior are. Here are some important examples:[7]

> Beliefs concerned with the moral wrongness of cheating.
>
> Beliefs and desires concerned with cheating in order to avoid personal failure, embarrassment, and so forth.
>
> Beliefs and desires concerned with not cheating in order to avoid getting caught, punished, and so forth.
>
> Beliefs and desires concerned with various costs (for oneself and others) of not following moral norms against cheating, such as potential social disapproval, guilt, lost trust, and so forth.
>
> Beliefs concerned with how to weigh these various costs and benefits.
>
> Desires concerned with not cheating when the benefits of cheating do not (significantly) outweigh the costs.
>
> Desires concerned with cheating when the benefits of cheating (significantly) outweigh the costs, while also desiring as much as possible to still be thought of as an honest person by oneself and others.

These are by no means all the beliefs and desires, which most of us have, that are directly relevant to cheating behavior, but they are central ones and help to illustrate my positive view.

Note that some of these mental states dispositions would, by themselves, earn a negative moral evaluation on most reasonable approaches to morality. Consider, for instance, the second and the last sets of beliefs and desires. Others, though, such as the first set seem to deserve a more positive evaluation. It is a good thing that most people seem to genuinely believe that cheating (at least in many familiar forms) is morally wrong.

Now consider the character trait that is made up of these and the other widespread mental state dispositions directly relevant to cheating.[8] My next claim is the following:

(5) This character trait, which most adult human beings possess, does not correspond to any of the words or concepts ordinary people have for traits associated with cheating.

---

[7] For relevant discussion, see Miller 2014: chapter 3.

[8] For the relationship between mental state dispositions and character traits, see my 2014: chapter 1. It is also important to note that there can be other, more narrow traits as well pertaining to certain dimensions of cheating, such as avoiding embarrassment. I am focusing at this broad level only for the sake of simplifying the discussion, but nothing about my view is incompatible with thinking that traits come in a wide variety of degrees of generality. For more, see my 2013: chapter 7.

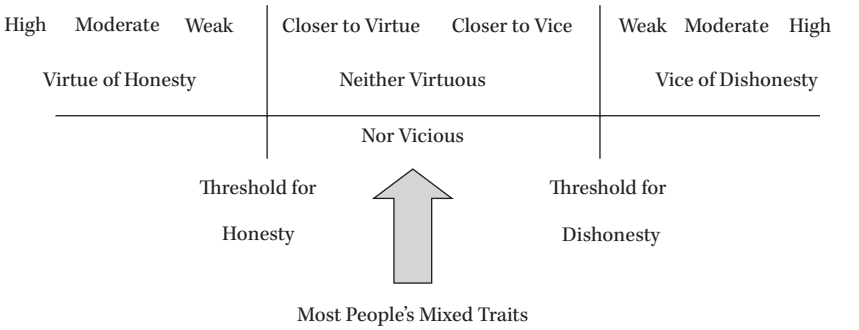

FIGURE 7.2 Where Most People Fall on the Continuum for Honesty and Dishonesty.

It is not the virtue of honesty as it fails to qualify as a moral virtue on a multitude of fronts. But it is also not the vice of dishonesty—it also fails to quality as a moral vice on a multitude of fronts.

So what is it? Well, if we do not have an ordinary term for it, then I have to invent one. I will call it a 'Mixed Cheating Trait', with the capital letters meant to serve as a reminder that this is a technical term that I have conjured up. The 'mixed' has to do solely with the moral evaluation of trait, as it has both morally positive and morally negative aspects. The 'cheating' has to do with the domain of morality to which it pertains. The story about why we never came up with a term for this trait and why we have seemed to overlook it for thousands of years of human history is an interesting one, and one that I have tried to tell a bit of in other work.[9]

So my view with respect to this moral domain can be put succinctly as follows:

(6) Most people have a Mixed Cheating Trait as part of their character, and not the virtue of honesty or the vice of dishonesty.

With this claim in mind, we can supplement figure 7.1 with figure 7.2. And, stepping back from the particular moral domain of cheating, my view can be stated in general terms as follows:

(7) Most people have a variety of Mixed Traits as part of their character and not a variety of virtues or vices.

This is compatible with it also being the case that:

(8) A few people have one or more virtues to some degree, rather than the corresponding Mixed Trait(s).

[9] See Miller 2014: chapter 7. It is important to be clear about the sense in which this trait is 'mixed.' The claim is *not* that the Mixed Cheating Trait is a virtue in some situations or contexts, and a vice in others. Rather, the claim is that this trait is *not a virtue in any situations or contexts*. Nor is it a vice in any situations or contexts. Nor is it entirely morally good or bad in any situations or contexts. Instead, a Mixed Cheating Trait is neither a virtue nor a vice, neither entirely good nor bad in *every* situation or context in which a person possesses it.

As well as with:

(9) A few people have one or more vices to some degree, rather than the corresponding Mixed Trait(s).

It is even compatible with:

(10) Some people do not have Mixed Traits, virtues, or vices when it comes to what causes their morally relevant behavior.

Such people might include infants, psychopaths, and those suffering from various severe mental handicaps, which do more than just block the activation of these character traits but actually undermine or prevent them from obtaining.

In the remainder of this section, let me highlight two additional aspects of Mixed Traits, one behavioral and the other motivational. We know that a moral virtue will reliably manifest itself in virtuous behavior when the person who has it is in relevant circumstances, other things being equal. We also know that a moral vice will reliably manifest itself in vicious behavior, other things being equal. A Mixed Trait, in contrast, should reliably manifest itself in . . . what?

Well if, according to figure 7.2, a particular Mixed Trait is intermediate between a virtue and a vice, then I think we should expect an individual who has it to display a variety of both morally positive and morally negative behaviors over time. Such a pattern of actual behavior can even appear fragmented to independent observers. Consider again the case of cheating, and recall the list of various beliefs and desires, which, I claimed, seem to be widely possessed. Then given such a collection, it should not be surprising at all to find a person who has them to be acting in diverse ways morally speaking. In some cases where there is an easy opportunity to cheat, a person with this Mixed Trait will do so. But in other situations, even ones varying to a slight degree, he often will not. For instance, the same person might cheat on a test when he knows he can get away with it, but in another class, when he has a chance to cheat undetected, he may not after having signed his school's honor code.[10] In the language of psychology, such a person exhibits a high degree of within-person variability in his cheating behavior.

So according to my framework:

(11) A person with a Mixed Cheating Trait will sometimes, without adequate moral justification, cheat when in certain situations in which opportunities to cheat present themselves, even without being forced or compelled to do so and even when the relevant rules are fair and appropriate.

But it is also true that:

[10] For the role of honor codes in curbing cheating behavior, see Miller 2014: chapter 3.

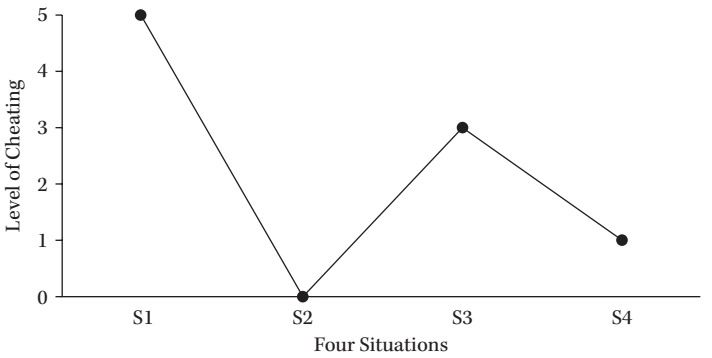

FIGURE 7.3 Larry's Cheating Profile (Peer Ratings).

(12) A person with a Mixed Cheating Trait will sometimes refrain from cheating in situations where he is a free and willing participant and the relevant rules are fair and appropriate, even if by cheating he is assured of acquiring some benefit for himself and knows he can get away with it.

Diagrammatically this can be illustrated as shown in figure 7.3, using an imaginary example of Larry's behavior in four different situations.[11] S1 could, for instance, be filling out his taxes, where he always cheats significantly; whereas S2 could be doing a take home exam for which he had to first sign an honor code, a situation where he never does anything wrong.

Figure 7.4 illustrates a different way to represent the same basic idea. Here we are seeing Larry's behavior in many different cheating situations over the course of an entire year. Note that, perhaps not surprisingly, much of the time he exhibits very little cheating. After all, most of us tend to not cheat, especially when others are around to notice, and there are obvious self-interested reasons for why that is the case. But even so, we can see that Larry exhibits the entire range of cheating behavior over the course of the year. There are plenty of situations where he cheats to a moderate degree (say, by inflating his charity donations by $50 dollars on his tax forms), and a few where he does so to a great extent (perhaps by looking up the answers online during a take-home exam). Thus it turns out that, given this within-person variability, how much Larry cheats during any given hour of this year does not reliably predict how much he cheats during any other given hour.[12]

[11] Here I am using a profile, a common tool in psychology, which in this case is a set of scores for the same kind of behavior exhibited by a person in different situations (Furr 2009).

[12] For similar claims about non-moral behavior, see Fleeson 2001.

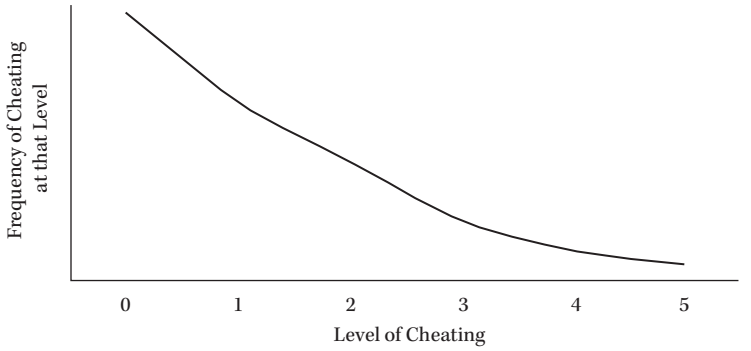

FIGURE 7.4 Larry's Behavior in Many Cheating Situations.

What holds for cheating and Mixed Cheating Traits also holds, to varying degrees perhaps, for the other domains of morality and the Mixed Traits pertaining to each of them. In other words:

(13) A person with a Mixed Trait will typically exhibit behavior from one situation relevant to that trait to the next, which can seem fragmented to observers.

This is a claim about overt behavior. But what can be said about Mixed Traits and the underlying motives behind that behavior?

Mixed Traits, like all moral character traits, affect behavior by first giving rise to occurent beliefs, desires, emotions, and the like. Those occurent mental states in turn causally issue in action. So what does the pattern of occurent beliefs, desires, and so forth we find with Mixed Traits look like on normative grounds? Here my answer is much the same as it was for actions. As traits intermediate between moral virtues and vices, these Mixed Traits are going to exhibit variable patterns of motivation—in some cases the motives will be admirable ones (such as helping another person out of empathetic concerns), whereas, in other cases, they will be deplorable ones (such as not helping another person so as to maintain a positive mood). There will be a high degree of within-person variability in the moral quality of the motivation, just as there was with the behavior.

Furthermore, as with moral traits in general, there will be psychological factors related to Mixed Traits which can increase motivation to perform certain actions, as well as psychological factors which can increase motivation to *not* perform other actions. Let me relate each of these points to the psychology of cheating.

In the case of a Mixed Cheating Trait, given the various beliefs and desires listed earlier, it is not surprising that the following are among the specific psychological variables which can trigger this trait and serve as *enhancers* for cheating behavior:

Anticipated lack of detection

Anticipated lack of punishment

Anticipated opportunity to avoid embarrassment

Anticipated opportunity to avoid failure

Note that these are not what we would consider to be virtuous enhancers for cheating—we would not expect an honest person's cheating behavior to fluctuate significantly depending on whether cheating would provide an opportunity to avoid embarrassment.

Stepping back, we can say more generally that:

(14) There are certain particular enhancers for a person's Mixed Cheating Trait which can influence the Trait in such a way that there is an increase in motivation to cheat when presented with opportunities to do so, as compared to the level of motivation to cheat apart from the influence of the enhancer.[13]

While it is no doubt overly simplistic, the role of an enhancer such as embarrassment avoidance can be illustrated using a diagram like the one in figure 7.5. Of course, it is artificial to suppose that Larry's cheating would increase to the *same* extent in these different situations, but the figure is only meant to help illustrate the basic idea.

Just as there are enhancers for cheating, so too:

(15) There are certain particular *inhibitors* for a person's Mixed Cheating Trait which can influence the Trait in such a way that there is an increase in motivation to not cheat when presented with opportunities to do so, as compared to the level of motivation to not cheat apart from the influence of the inhibitor.[14]

In this case examples include:

Anticipated cheating failure

Anticipated guilt

Psychologically salient moral norms against cheating

These should be largely intuitive, and there is support in the psychological literature for each of them functioning as inhibitors.[15]

---

[13] In one sense, (14) will be true for any character trait, as there will always be various factors which can lead to increased motivation to act a certain way. The intended emphasis, though, is on *certain particular* enhancers which are tied to a Mixed Cheating Trait, and which are not necessarily enhancers for other traits such as honesty. Also, note that the psychological activity involved with these enhancers need not be carried out in a conscious manner—often it is largely automatic or is at work at the subconscious level.

[14] Similar remarks as in the previous footnote would apply to (15) as well.

[15] For more on enhancers and inhibitors, see Miller 2013: chapters 7 and 8. It is important to stress, though, that the claim is *not* that inhibitors, when active, simply lead to the same levels of cheating

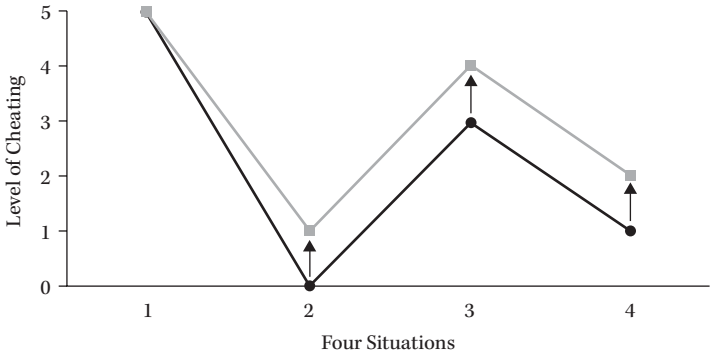

FIGURE 7.5 The Influence of an Enhancer on Larry's Cheating Profile (Peer Ratings).

To summarize, the more general claim about motivation and Mixed Traits is the following:

(16) For each Mixed Trait, we should expect a high degree of within-personal variability in the moral quality of the motives to which it gives rise. Furthermore, certain enhancers and inhibitors can influence motivation to act in trait-relevant ways relative to what those levels would be were they not present. What the enhancers and inhibitors are for any given Mixed Trait, is a function of the mental state dispositions which make up that trait.

Clearly no simply motivational story is going to be available for the various moral domains if my approach is correct.

Other features of Mixed Traits could be clarified—they are stable and cross-situationally consistent in certain ways, they can differ from one person to the next, they vary in how specific or general they can be individuated, they often carry out their psychological processing without our conscious awareness, and so forth. But I will pass over these details for now.[16]

2. The Moral Domain of Resource Distribution

As noted earlier, I have explored the Mixed Trait framework extensively with respect to some moral domains, but certainly not all of them. In the remainder of this chapter I want to touch on two additional domains.

---

motivation and behavior which would typically be found in control participants if we were doing studies of cheating. Rather, the claim is that they give rise to *motivation to not cheat* in a certain situation. So inhibitors, when the right conditions obtain, are expected to reduce cheating rates *below* those seen in controls.

[16] See, in particular, Miller 2013: chapters 7 and 8.

One moral domain in which there has been extensive psychological research is that of resource distribution. There is no way to engage with all of this literature, so let me focus on a recent series of influential and important studies by Daniel Batson and his colleagues. Suppose you are in a situation in which you have to make a choice between acquiring something that is only moderately conducive to your own self-interest versus making sure that the same thing goes to another person instead. The choice is real, not hypothetical, and we can further stipulate that it is simple and mundane—plus, let me add that there is clearly widespread consensus that the morally correct thing to do in this case is to choose in favor of the other person's interests. We can even make this fairness principle highly salient to you beforehand (you knew it already going into the situation, but it can't hurt to call it to mind again), and we can follow up by asking afterward whether you believed it or not. Suppose we do all this, and you do report that you believe the moral principle. Then it looks like a perfect kind of case in which you would reliably help out the other person. Yet in study after study with this setup, Batson found that most people help themselves, not the other person.[17]

Here is the setup Batson typically used. Participants were told that they were part of a task assignment study. Individually, they were given the choice of whether to assign a positive consequences task or a neutral consequences task to either themselves or another participant, who (they are told) will simply assume the assignment was made by chance. The positive-consequences task was such that for each correct response, the participant would receive one ticket for a raffle with a prize of $30 at the store of his or her choice. In the neutral consequences task, there would be no consequences for correct or incorrect responses but, "most participants assigned to the neutral consequences task find it rather dull and boring."[18] After making the assignment privately and anonymously, participants were asked about what was the morally right way to assign the task consequences, and to rate on a 9-point scale whether they thought the way they had actually made the task assignment was morally right.

What would you have done if you had been given this task? Well, if you are like most people, you would have assigned yourself the positive task. Out of 20 participants,[19]

---

[17] For the studies, see Batson et al. 1997, 1999, 2002, 2003. For reviews, see Batson and Thompson 2001; and Batson 2008. For related studies and discussion, see Valdesolo and DeSteno 2007, 2008; and Watson and Sheikh 2008. For criticism, see Fernandez-Dols et al. 2010.

I discussed Batson's research in my 2013: chapter 4, in connection with the virtue of compassion. But I have since thought that it is more relevant to the virtue of fairness, a connection which will be explored in this section.

[18] Batson et al. 1997: 1339.

[19] Ibid., 1340.

Assigned self to positive consequences task: 16
Assigned other to positive consequences task: 4

Furthermore, only 1 out of the 16 said that assigning oneself to the positive task was morally correct. Yet even these 16 participants rated the morality of their assignment in the middle of the 9-point scale (4.38).[20]

From here Batson performed many additional studies which were variations of the initial setup. Suppose, for instance, that we make the moral norm at work here salient to participants just before they make their assignment by including the statement: "Most participants feel that giving both people an equal chance—by, for example, flipping a coin—is the fairest way to assign themselves and the other participant to the tasks (we have provided a coin for you to flip if you wish). But the decision is entirely up to you."[21] Ten participants flipped and 10 did not. Eight out of 10 in the first group said flipping was the morally right procedure, and 6 out of 10 said so in the second group. Most importantly,[22]

Assigned self to positive consequences task out of 10 who did not flip: 9
Assigned self to positive consequences task out of 10 who *did* flip: 9

This second result is grossly out of line with what the random flipping of a coin would have predicted. At least some of the participants in the second group must have flipped in a way that went in favor of the *other* person but still assigned themselves the positive task. Self-interest seemed to have crept into their decision-making process in a significant way. Yet, and perhaps most surprisingly, those who flipped rated what they had done much more morally right (7.30 on a 1–9 scale) than those who did not flip (4.00) did.[23]

Here is another wrinkle—suppose participants are now given the option to have the experimenter assign them one of the two tasks , while also knowing what the assignment is going to be ahead of time. We get the following:[24]

Accepted experimenter's assignment if it was going to be the positive consequences task 17 out of 20
Accepted experimenter's assignment if it was going to be the neutral consequences task 11 out of 20
Remaining participants who assigned themselves to the positive consequences task (whether flipped or not) 12 out of 12

[20] Ibid. Although that was significantly lower than the 8.25 rating for the 4 participants who assigned the other participant to the positive consequences task (ibid.).
[21] Ibid., 1341.
[22] Ibid., 1342.
[23] Ibid., 1341.
[24] Ibid., 1343.

In this experiment, only 22.5% of participants ended up with the neutral task. Yet for the 17 participants in the first group, on average they felt they had acted just as morally right (7.06 on a 9-point scale) as the 11 participants in the second group (7.91).[25]

What these results have suggested to Batson is that most of us (at least in these kinds of situations) are disposed toward a kind of *moral hypocrisy*, or appearing to be moral to oneself and others but avoiding the costs of actually being so if we can (try to) get away with it. After all, many participants were typically eager to flip the coin and report after the fact that this was the morally right course of action, but then distorted the process so that the results came out in their favor. Note that it is not the mere fact that they choose the positive consequences task for themselves that is hypocritical by itself—there could be cases where a person thinks of an act like this as being in line with her self-interest, and is not even aware of its moral ramifications. Nor is the hypocrisy captured here by the additional fact that the participants also seemed to believe in the moral principle that flipping the coin is the morally fair way to assign the task consequences. For then we just have a perfectly familiar case of weakness of will, in which you believe that something is right but fail to be sufficiently motivated to do it. Rather, the hypocrisy arises when they (i) choose the positive consequences for themselves, while (ii) seeming to believe that flipping the coin and following what it indicates is morally correct, *and* (iii) still claiming to themselves and others to have made the morally right task assignment.[26]

How is it really possible for these participants to pull off this combination? In particular, it is one thing to appear to be acting morally to others. But these participants are appearing to be moral *to themselves too*. How are they able to downplay the costs associated with guilt, regret, and hypocritical behavior for acting immorally by going against what they know to be right, and also experience the self-rewards for moral behavior? One possibility is that the participants in question have come to deny any moral responsibility in this situation, thereby deactivating the moral norm. Or perhaps they have come to deny that they are able to carry out the task, or that they are no longer aware of the likely consequences of their actions. But there is no evidence to suggest that these background conditions have failed to apply in these particular cases.[27] Another possibility is that these participants have come to think that their behavior is in line with their moral standards, thereby allowing them to not feel guilt and to even perhaps take pride in their behavior. This could be because they have revised the content of the moral principle to create an exception clause for

[25] Ibid.

[26] For more on characterizing moral hypocrisy, see Batson et al. 1997: 1335–1336; 1999: 525–526; 2002: 330; Batson and Thompson 2001; and Batson 2008; 2011: 222–224.

[27] See in particular Batson 2008: 57. Studies involving other types of situations have found denial of responsibility, for instance, to be an important part of the explanation for a failure to conform to personal norms. See, e.g., Schwartz 1973; 1977: 230, 256–262.

this one kind of situation. Or it might be that they lie to or deceive themselves about what the principle says in the first place.

But the results already cited cast doubt on these hypotheses as well.[28] Batson also introduced a variation of the setup whereby the moral principle is made salient, a coin is provided, and the coin is clearly marked on one side "SELF to POS[ITIVE]", and "OTHER to POS" on the opposite side.[29] In those cases when the coin lands on OTHER, it seems very hard to think that a person who still assigns himself to the positive consequences task—in spite of what the moral principle and the coin both say—would take that to be morally acceptable. Furthermore, only 2 out of 40 participants reported afterward that the most morally right thing to do is to assign themselves to the positive consequences task.[30] Yet of the 28 who chose to flip:[31]

Assigned self to positive consequences task	24
Assigned other to positive consequences task	4

Furthermore, those who assigned themselves to the positive consequences task after flipping the coin again thought that they were being highly moral (7.42 on a 9-point scale), while those who made the same assignment without flipping the coin did not (3.90).[32] So marking the coin and thereby reducing the ambiguity as to what the fair assignment should have been, did nothing to undermine moral hypocrisy. We are still without an explanation for how it seems to work—more on that in a moment.

---

[28] Batson et al. 1999: 526. This is not to say that these hypotheses are not accurate in other cases, but the focus here is only on understanding the data generated by the Batson's studies.

[29] Ibid., 527.

[30] Ibid., 529.

[31] Ibid., 528.

[32] Ibid., 529. A natural thought here is that this result does not distinguish between those participants who flip, win, and then rate the morality of their action versus those who flip, lose, change the task assignment to favor themselves, and then rate the morality of their action. It could be that the former group (rightly) rates the morality of their action highly, since they are following the fair procedure, whereas the second group rates it low, perhaps around 4.0, as do those who do not even bother flipping the coin and just assign themselves the positive task. If the results came out this way, then we would have evidence for moral weakness of will, to be sure, but not for moral hypocrisy, since the second group is honest about the moral failure of their action.

In a later study, Batson tested this possibility using secret observations of how each participant who flipped ended up having his or her coin fall (2002: 334–338). Of the 32 who flipped the coin, 16 got OTHER to POS or otherwise fiddled with the coin flip so that it came out in their favor (i.e., repeated flipping). For this group, the mean rated morality of their action was 5.56, which was significantly higher than the 3.89 for the group which assigned itself to the positive task without using the coin. On the other hand, it was also lower than the 7.45 for the group who flipped, won SELF to POS, and assigned themselves the positive task (p. 336). The upshot is that, "even though the coin had no more effect on their decision than it had on the decision of those who did not claim to use the coin at all, the fiddlers still said they thought the way they made the task assignment decision was more moral. Their sham reference to use of the coin seems to have provided sufficient appearance of morality that they could claim to have acted, if not totally morally, at least moderately so" (337).

To summarize, the studies mentioned above suggest that participants could have had their relevant fairness norm activated—in this case, the norm that flipping a coin and conforming to its results is the morally right way to behave—and even be motivated to some extent to do so. But this motivation must have been so weak that, once the cost/benefit analysis was done on an alternative action of secretly ignoring the coin when it went against the person and instead assigning oneself to the positive task, it became fairly easy to outweigh or undercut the motivation to do the right thing. So our fairness norms look to have little enhancing power after all, or at least in these kinds of cases.

But all is not lost for fairness norms. Batson was able to discover two variables which separately helped to ensure that motivation to be fair did win out. The first variable was the introduction of a mirror. The setup was as usual with no labeling of the coin, but now a mirror was propped up against a wall on the only table where participants could fill out their forms. For one group of participants, the mirror was facing them; for the other, it was turned to the wall. This ended up making a significant difference:[33]

	Assigned Self to Positive Consequences Task	
	Not Facing Mirror	Facing Mirror
Did not flip coin (29)	85%	62%
Flipped coin (23)	85%	50%

Here for the first time we see results in line with what fairness requires—for those who flipped the coin and were facing the mirror, the overall results were as chance would predict.

What is the best explanation for the contribution made by the mirror? Batson appeals to research on the psychology of self-awareness, in which objects like mirrors can, "heighten awareness of discrepancies between behavior and salient personal standards, creating pressure to act in accord with standards."[34] In this case, the mirror served to highlight to the person the difference between what he believed was morally correct in this situation, and opposing temptation to act out of self-interest. This heightened awareness seemed to either create extra motivation to comply with the fairness norm or lessen motivation to do the self-interested thing (or perhaps both).[35]

Self-awareness also provides a clue about where to find a plausible explanation for how moral hypocrisy is possible. That clue has to do with a particular

---

[33] Batson et al. 1999: 530.

[34] Batson et al. 1999: 529. For more on the psychology of self-awareness, see Wicklund 1975.

[35] Thus the mirror can be said to increase the salience of the self's own personal standards of evaluations. Another hypothesis, though, is that it increases the salience of standards of social evaluation, i.e., how others might judge him. Batson tested this possibility and did not find support for it (2002: 331–334).

form of self-deception.[36] Rather than thinking that participants simply revised their fairness standards to make their behavior look acceptable in their own eyes, perhaps instead they were engaging (often unconsciously) in an act of self-deception whereby they avoided comparing their behavior to the relevant moral standards. If the two are kept apart from each other, that mitigates the perceived costs of not acting fairly while doing nothing to mitigate the perceived benefits of acting self-interestedly. But with increased self-awareness, the discrepancy between the fairness norm and the self-interested option was made especially salient so that it became psychologically difficult for many participants to employ this particular form of self-deception.[37]

The other variable which Batson found to increase motivation associated with the fairness norms had to do with perspective taking. Take the usual setup, but with the caveat that the default starting point is for the participant to be awarded two raffle tickets for every correct response, while the other participant gets zero tickets. Then the task assignment becomes whether the participant is willing to change the assignment so that it is symmetrical with both people receiving one ticket each. For controls, 38% changed the task consequences to symmetrical. However, for the experimental group, 83% did.[38] The difference? This group was instructed to adopt the perspective of the other person—"we would like for you to imagine yourself in the place of the other participant."[39] This is different from imagining what the other person is feeling or experiencing—the instructions here are very much in line with the Biblical mandate to, "Do unto others as you would have them do unto you."[40]

There is one additional implication from Batson's work which I want to highlight here. Consider again those participants who are about to flip the coin. At that point before they see what the outcome is, what is the nature of their motivation? Perhaps for the moment at least some of them want to do the morally right thing (follow the dictate of the coin, however it ends up landing) for its own sake. If the coin lands in their favor, then the outcome also aligns with their self-interest, which is so much the better. But if it lands in the other person's favor, then they might see what being fair would cost them, and

---

[36] See Batson et al. 1997: 1336, 1346; 1999: 526–527.

[37] Batson et al. 1999: 527, 529, 531–532; 2002: 331; and Batson and Thompson 2001: 55. Batson et al. 1999 ran a third study in which they examined whether the motivation involved in self-awareness was still moral hypocrisy as opposed to a motive to care about morality for its own sake, and found evidence in favor of the first option. For very interesting discussion of some of the morally problematic effects of increased self-awareness, see Batson et al. 1999: 535.

[38] Batson et al. 2003: 1199.

[39] Ibid., 1198; emphasis removed.

[40] Matthew 7:12. For more on the difference between these two perspectives, see Miller 2013: chapter 5.

Interestingly, in the familiar variant of the experiment with a positive and a neutral task assignment, participants who adopted this perspective did not assign the other person to the positive task any more than did participants who did not do any perspective taking (25% in each case; ibid.: 1195). For an explanation of the different results in the two experiments involving perspective taking, see Batson et al. 2003: 1199–1200.

their self-interested motives end up outweighing their initial moral motivation. Or perhaps this is all fanciful—perhaps when they are flipping the coin, all they want is what they think is ultimately in their self-interest, which at the moment is just flipping the coin so as to benefit from appearing to be moral.[41]

One way to test these hypotheses, Batson reasoned, is to see if participants cared about whether the flipping of the coin and so the task assignment was done by themselves or by the experimenter. If the egoistic hypothesis is correct, then they should want to flip the coin themselves so they can rig the outcome. If the other hypothesis involving an ultimate desire to be moral is correct, however, it should not matter who flips. When the experiment was actually run with this choice option, 80% of those who used a coin wanted the experimenter to flip it. This initial evidence thus favors postulating a motive to follow the fairness norm (and perhaps moral norms more generally) for its own sake.[42]

Suppose this hypothesis about motivation is correct. How strong and psychologically powerful a force does it typically seem to be in cases involving fairness? The evidence suggests, at least given the current state of research, that for many of us it is only weak in strength. We can already see this from the studies cited earlier in which so many participants do not actually follow their moral principle about what a fair task assignment is. In addition, Batson varied the previous setup so that the assignment was between a positive and a *negative* task, where the latter involved receiving "mild but uncomfortable" electric shocks for each incorrect response. With this change, only 25% of participants offered to let the experimenter flip the coin, and another 25% flipped themselves, with 91% choosing the positive task. The remaining 50% of participants simply bypassed the pretense of the coin flip and gave themselves the positive task while readily admitting that this was not morally right.[43] The implication is that moral motivation caused by our fairness norms seems to be weak and highly susceptible to being outweighed in cases of this type in which the person's self-interest is at stake.

Having reviewed this body of psychological research, let me step back and consider some of its wider implications for our understanding of people's moral characters. Recall that I have advanced both a negative claim about the lack of widespread virtue and vice, together with a positive claim about Mixed Traits.

---

[41] More generally, Schwartz suggests that we have a "desire to act in ways consistent with one's values so as to enhance or preserve one's sense of self-worth and avoid self-concept distress" (1977: 226). Such proposals naturally call to mind Kant's claim: "Out of charity I am willing to grant that most of our actions are in accord with duty; but if we look more closely at the devising and striving that lies behind them, then everywhere we run into the dear self which is always there; and it is this and not the strict command of duty (which would often require self-denial) that underlies our intentions" (2002, 4:407).

[42] See Batson and Thompson 2001: 55–56.

[43] Ibid., 56.

Begin with the lack of widespread virtue. In this context, the virtue in question would be fairness (or justice more generally), and it seems clear enough that Batson's results do not inspire much confidence in that virtue being widely held. For instance, something like the following seems plausible to me:

(17) A person who is fair will regularly distribute a good fairly as opposed to selfishly, especially when he or she believes that this would be the fair thing to do and the good is relatively small in value.

But time and again in Batson's studies, the participants rigged things so that they received the better outcome.

Similarly, I would expect this to be true of the virtue of fairness:

(18) A person who is fair will not regularly claim to have done the morally correct thing when he knows that what he did was unfair and selfish and that the benefit to himself was relatively small in value.

Again, though, the participants did not tend to live up to this standard. As we saw, in one study those who flipped and assigned themselves the positive consequence task rated what they had done as highly moral (7.30 on a 1–9 scale), and their rating was much higher than for those who assigned themselves this task without flipping (4.00).[44]

There are other respects in which participants failed to live up to the standards of fairness, but let me switch now to the vice of unfairness, and note that the picture is not all bleak. In fact, there are many respects in which participants did not conform to the standards of unfairness either. Here is one:

(19) An unfair person does not have moral beliefs to the effect that unfairness is wrong in general, as well as wrong in particular instances of what are widely considered to be acts of unfairness. Or, if he does happen to have such beliefs, he will not care much about them and they will not play a significant motivational role in his psychology.

But we saw that Batson's participants did seem to genuinely believe that the fair thing to do was to flip the coin and make the appropriate assignment. True, this belief did not have a significant motivational role to play in some contexts, but in others it did, such as when taking the perspective of the other person.

This suggests another respect in which most people, if they are like Batson's participants, would fall short of being vicious in this context:

(20) An unfair person would not let his unfair behavior be significantly diminished or eliminated given either increased self-awareness or when momentarily adopting the perspective of another person.

---

[44] Batson et al. 1997: 1341.

Yet the experimental manipulations Batson introduced involving a mirror and the instructions to imagine yourself in the place of the other person, seemed to eliminate the unfair behavior altogether.

To note one more interesting conflict with unfairness, consider the motivational picture which emerged in the last few studies that were reviewed concerning fairness norms. Batson's results suggested that participants were motivated to do the morally right/fair thing, but when the flip of the coin went against them, this dutiful motivation was often outweighed by self-interested motivation. Yet on a traditional Aristotelian picture of the vices,

(21) An unfair person is not motivated to do the morally right thing, even if that motivation ends up being outweighed, but rather is motivated to do that which happens to be unfair.

Someone who is motivationally conflicted but gives into self-interest is weak of will or incontinent on the Aristotelian picture, rather than being vicious.

So, while each of these points could be expanded and other incompatibilities noted, it seems apparent to me that Batson's results do not sit comfortably with either the widespread possession of fairness or of unfairness. Instead, they suggest that we have beliefs and desires like these:

Beliefs concerns with what is fair and what is not fair, and the importance of being fair.

Beliefs concerned with the relationship between not complying with certain fairness norms and various personal benefits, such as increased time, money, alternative activities, and so forth.

Desires concerned with being fair when doing so will contribute toward complying with the relevant fairness norms, provided the benefits of doing so are not (significantly) outweighed by the costs.

Desires concerned with not being fair when the benefits of complying with the relevant fairness norms are (significantly) outweighed by the costs, while also desiring to as much as possible still appear to be fair both to others and to oneself.

While this is certainly not an exhaustive list, these mental states do indeed sit very comfortably with my Mixed Trait approach. Some of them are quite morally admirable, such as those in the first set, and by themselves could give rise to positive moral behavior. Others, of course, are not morally admirable, and they can help to explain the unfair behavior we see in Batson's studies.

Hence in my view, Batson's findings lead naturally to my Mixed Trait approach, and give me good reason to postulate the widespread possession of a 'Mixed Distribution Trait.' Again this trait is 'mixed' solely on normative

grounds—it is not good enough to qualify as a virtue but not bad enough to qualify as a vice. And the 'distribution' is meant to signify the moral domain in question having to do with the distribution of resources. Finally, this is very much a trait—it is a disposition to think, feel, and act in certain ways, which consists (in part) of the beliefs and desires listed here.

What about the two aspects of the Mixed Trait approach that were highlighted in section 1—how do they measure up here? The first is that we should expect a high degree of within-person variability in fair and unfair behavior. Now like almost all studies of moral behavior in psychology, Batson's experiments did not track the same participants longitudinally in multiple situations. However, it is reasonable to expect that the patterns he found using different participants in different situations would also hold for the same participants in these situations. For instance, they would likely assign themselves the positive task in the standard setup, but a month later, would likely make the assignment fairly if instructed to take the perspective of the other person. So it seems to me that there is every reason to expect a wide variation in fairness behavior, even when only minor changes are made to the situations, such as the presence or absence of a mirror.

The same is true for motivation as well. In some cases, people might be motivated to do the fair or moral thing, and for its own sake. Furthermore, that motivation might be strong enough to lead to fair action. But we also saw plenty of instances where egoistic motivation got the upper hand in ways that were patently unfair. So there is good reason to expect that the second aspect of my Mixed Trait approach is also being exemplified—a high degree of within-person variability in the quality of a person's moral motivation.

Now it is worth stressing that I have only examined one area of the research literature on resource distribution, and clearly much more needs to be said. But so far I am encouraged by the results as far as my model is concerned.

3. The Moral Domain of Theft

In comparison to section 2, my discussion of theft in this section will be very brief. As we shall see, the reason is not because I am trying to downplay or neglect some difficult experimental results for my view; rather it is simply because there are almost no published studies in this area that can shed much light on our moral characters.

In the case of theft, this is perhaps not surprising. The best kind of study for examining moral character is a behavioral study, ideally in a natural environment in which people do not even know that they are being observed. Or, absent such a setup, the best studies should at least involve a laboratory environment in which one or more variables are manipulated to see what impact

they have on subsequent behavior in comparison to a control group.[45] But now consider trying to run either kind of study when the behavior in question is theft—that naturally introduces a host of practical and ethical challenges. So in contrast to other moral domains, such as resource distribution or cheating, it is not surprising that behavioral research on theft is rare at best.

In fact, I am not familiar with *any* experimental studies published in leading psychology journals which are primarily focused on theft behavior in adults.[46] Instead, what we find are self-report studies, which for a variety of reasons shed much less light on what our moral characters are actually like.[47] A 1989 study by Karen Slora offers a nice illustration. She collected 341 usable surveys of employees in the fast-food industry, and some of the notable results included:[48]

	% Admitting
Eating food without paying	54
Taking money without permission	6
Employee discount used for friends	30
Faking injury for workmen's compensation	1

Furthermore, each of these behaviors was broken down into self-reported frequency of behavior. In the case of taking money without permission, for instance, almost everyone reported doing this only once in a six-month period.[49] Furthermore, the amount of money taken was almost always said to be $5 or less per week on average.

Slora also ran a similar survey for the supermarket industry, and collected 281 usable surveys. Here are the results for the same questions:[50]

	% Admitting
Eating food without paying	35
Taking money without permission	1
Employee discount used for friends	11
Faking injury for workmen's compensation	0

And both the frequency and amount of money taken were even lower than for the fast-food participants.

---

[45] For defense of these claims, see Miller 2013: chapter 8.

[46] For a study involving children and Halloween candy, see Diener et al. 1976. For children and stealing gift certificates, see White 1972. There are also well known experimental studies of bystander interventions to stop a theft (see, e.g., Moriarty 1975), but that is a very different issue from experimental studies of theft behavior itself.

[47] For extensive discussion, see Miller 2013: chapter 8.

[48] Slora 1989: 204–205.

[49] Ibid., 204.

[50] Ibid., 211–212.

No doubt this is important data to collect for a variety of purposes, but it should be obvious enough that it sheds little light on the participants' level of honesty or dishonesty when it comes to stealing. First of all, there are the usual concerns with self-report data, especially when some of the behavior in question is very morally significant and could lead to severe disciplinary action. As Dalton and Metzger note, even with assurances given to participants that their anonymity would be protected, "one would need an exceptional confidence in the scientific community to admit complicity in some stigmatizing or illegal act under these circumstances."[51] But even leaving aside these concerns, we do not have much of a sense of *how many* opportunities there even were for these employees to actually engage in behaviors like taking money unobserved. Perhaps many of them would have done so if given the chance, but the relevant opportunities never (or only rarely) emerged. Furthermore, even if there *were* such opportunities, we also do not know to what extent these employees knew they were available or not. They could have not stolen, only because they failed to notice that they had a good chance to do so. Finally, it is also important to stress that this work does not shed any light on the underlying motives that were present for stealing or not stealing. Even if someone refrained from stealing from the cash register when no one was looking, that would not be a reflection of a virtue per se if it was done primarily to, say, avoid feeling guilty afterward.

So studies like Slora's have some inherent limitations in terms of how well suited they are to revealing anything about these employees' character. As a result, they do little to adjudicate between competing empirical accounts of what our moral character looks like, including whether most of us are honest, dishonest, or have Mixed Stealing Traits. Indeed, one could imagine someone citing the low levels of self-reported theft of money as evidence for the virtue of honesty. Someone else could look at Slora's results and cite the high levels of eating food without paying for it as evidence of dishonesty. Or, my model could comfortably accept both of these findings. But it is not clear that anything in these results really tells in favor or against any of these three options.[52]

More recently, researchers have devised clever ways to try to address some of the self-report worries. For instance, James Wimbush and Dan Dalton surveyed three groups of employees at "high-theft-exposure businesses," such as fast-food, convenience, and service stores. One group received a conventional theft survey focused specifically on stealing cash, supplies, or merchandise. Another received the same questions but using a RRT protocol, which guaranteed that an experimenter would not be able to tell if any particular respondent

[51] 1992: 209. See also the related discussion in Wimbush and Dalton 1997: 757.

[52] For additional self-report studies on employee theft behavior, see Jones et al. 1990; Boye and Slora 1993; and Boye and Wasserman 1996, as well as the review in Wimbush and Dalton 1997.

had admitted to a theft behavior. And the third group completed a survey using an UCT protocol, which works by having respondents give aggregate responses to a group of questions rather than answer any particular question.[53] The thought was that either of the latter two approaches would do a better job of making participants feel comfortable in accurately reporting their real theft behavior.

Compared to the first survey group, one might have expected self-reported theft to be higher in the second and third groups, especially if one doubts that most people are honest. And that prediction was indeed supported:[54]

	Conventional Survey	RRT Protocol	UCT Protocol
% Admitting theft	28.2	59.2	57.9
% Admitting theft (once per month) at different theft levels:			
$5.00–$9.99	12.9	26.1	21.0
$10.00–$24.99	8.6	17.2	16.2
$25.00–$49.99	4.3	10.4	13.2
> $50	2.4	5.5	6.7

One thing that jumps off the page is the disparity between the results of the conventional survey and these other two techniques, which suggests that the former (the dominant approach in the research literature) may be badly underrepresenting actual theft behavior.

What should we make of these results? Again, at least as far as assessing moral character is concerned, the limitations I noted above apply here too, including basic worries about self-reporting, since even these improved methods do not help address either simple forgetfulness or more subtle self-deception or rationalization of past behavior as not constituting theft. Furthermore, most of the self-reported theft here is very minimal in monetary amount, and could just amount to a soda and chips once a month.[55] So it is hard to know whether such behavior would really count against someone's being honest or not. And even if the results do put some pressure on the widespread possession of honesty, they would not do anything to differentiate between a Mixed Stealing Trait and the vice of dishonesty.

Let me end by looking at one more issue in this literature. Several studies have tried to identify variables which are correlated with either increased or decreased levels of employee cheating. The work of John Kamp and Paul

[53] These are oversimplifications of the protocols. For more details about both the RRT and UCT protocols, see Wimbush and Dalton 1997: 758.

[54] Ibid., 760.

[55] See Ibid.

Brooks is one such example.[56] Using 132 usable surveys from college-student employees, they found that the following were significantly correlated with theft (including discount abuse, merchandise theft, and cash theft):[57]

	Correlation
Company policies	.23
Supervisor control	.27
Coworker attitudes	.64
Coworker theft	.52
Personal attitudes	.49
Job dissatisfaction	.36

The correlations with coworker attitudes and coworker theft are especially notable given how high they are.

It is tempting to use these particular results to argue against the honesty of most of these employees. After all, an honest person would not steal, regardless of whether or not his or her coworkers were stealing. Nor, in cases where her coworkers have accepting attitudes toward stealing from their employer, would an honest person let those attitudes significantly sway her toward stealing as well. So, given this data from Kamp and Brooks, we might doubt the general honesty of these college students.

I do agree that these are reasonable expectations for an honest person. But unfortunately, so far as I am aware we do not have any published data in leading journals to use to adequately test them. In the case of this particular study, one serious limitation is that Kamp and Brooks are reporting only correlations, and so their results do not tell us whether the coworker attitudes or behavior actually did anything to causally influence the theft.[58] And furthermore, the data they collected is, once again, merely self-reported.

So as far as the moral domain of stealing is concerned, then, my conclusion is that the current state of the research literature gives me no reason to worry about any incompatibilities with my Mixed Trait model. But on the flip side, it also gives me no basis upon which to evaluate whether most people have the virtue of honesty, the vice of dishonesty, a Mixed Stealing Trait, or perhaps some fourth kind of trait. Hence the real conclusion of this section is that we are badly in need of new and innovative research which probes this area of our moral lives.

---

[56] See Kamp and Brooks 1991. For additional work, see Hollinger and Clark 1983; Bolin and Heatherly 2001; and Kulas et al. 2007.

[57] Kamp and Brooks 1991: 453.

[58] As the authors themselves note. See ibid., 456–457, for discussion.

4. Conclusion

This chapter has ended up telling a sharply contrasting story about what the experimental research literature allows us to say concerning our character in two domains of our moral lives. On the one hand, very carefully constructed behavioral studies have been done which probe aspects of our motivation and behavior when it comes to the fair distribution of a good. And these studies, in my view, do end up offering additional support for my Mixed Trait model as well as additional reasons to doubt that most people have the relevant virtue or vice. On the other hand, there are almost no behavioral studies of stealing behavior, and the self-report studies we do have tell us very little about what we are really like in this area. This is not bad news for my model, but it is not good news either. It is best understood as a call for patience and an expression of a hope that psychologists will increasingly turn their attention to this badly neglected area of study.[59]

References

Alfano, Mark. (2013). *Character as Moral Fiction*. Cambridge: Cambridge University Press.

Adams, Robert. (2006). *A Theory of Virtue: Excellence in Being for the Good*. Oxford: Clarendon Press.

Batson, C. (2008). "Moral Masquerades: Experimental Exploration of the Nature of Moral Motivation." *Phenomenology and the Cognitive Sciences* 7: 51–66.

———. (2011). *Altruism in Humans*. New York: Oxford University Press.

Batson, C., D. Kobrynowicz, J. Dinnerstein, H. Kampf, and A. Wilson. (1997). "In a Very Different Voice: Unmasking Moral Hypocrisy." *Journal of Personality and Social Psychology* 72: 1335–1348.

Batson, C., D. Lishner, A. Carpenter, L. Dulin, S. Harjusola-Webb, E. Stocks, S. Gale, O. Hassan, and B. Sampat. (2003). "'. . . As You Would Have Them Do Unto You': Does Imagining Yourself in the Other's Place Stimulate Moral Action?" *Personality and Social Psychology Bulletin* 29: 1190–1201.

Batson, C., and E. Thompson. (2001). "Why Don't Moral People Act Morally? Motivational Considerations." *Current Directions in Psychological Science* 10: 54–57.

Batson, C., E. Thompson, and H. Chen. (2002). "Moral Hypocrisy: Addressing Some Alternatives." *Journal of Personality and Social Psychology* 83: 330–339.

Batson, C., E. Thompson, G. Seuferling, H. Whitney, and J. Strongman. (1999). "Moral Hypocrisy: Appearing Moral to Oneself without Being So." *Journal of Personality and Social Psychology* 77: 525–537.

---

[59] Section 2 of this chapter draws on material from Miller 2013: chapter 4. Support for working on this chapter was funded in part by a grant from the John Templeton Foundation and by a grant from the Templeton World Charity Foundation. The opinions expressed here are my own and do not necessarily reflect the views of these foundations.

Bolin, A., and L. Heatherly. (2001). "Predictors of Employee Deviance: The Relationship between Bad Attitudes and Bad Behavior." *Journal of Business and Psychology* 15: 405–418.

Boye, M., and K. Slora. (1993). "The Severity and Prevalence of Deviant Employee Activity within Supermarkets." *Journal of Business and Psychology* 8: 245–253.

Boye, M., and A. Wasserman. (1996). "Predicting Counterproductivity among Drug Store Applicants." *Journal of Business and Psychology* 10: 337–349.

Diener, E., S. Fraser, A. Beaman, and R. Kelem. (1976). "Effects of Deindividuation Variables on Stealing among Halloween Trick-or-Treaters." *Journal of Personality and Social Psychology* 33: 178–183.

Doris, John. (1998). "Persons, Situations, and Virtue Ethics." *Noûs* 32: 504–530.

———. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Fernandez-Dols, J., P. Aguilar, S. Campo, R. Vallacher, A. Janowsky, H. Rabbia, S. Brussino, and M. Lerner. (2010). "Hypocrites or Maligned Cooperative Participants? Experimenter-Induced Normative Conflict in Zero-Sum Situations." *Journal of Experimental Social Psychology* 46: 525–530.

Fleeson, W. (2001). "Toward a Structure- and Process-Integrated View of Personality: Traits as Density Distributions of States." *Journal of Personality and Social Psychology* 80: 1011–1027.

Furr, R. (2009). "Profile Analysis in Person-Situation Integration." *Journal of Research in Personality* 43: 196–207.

Harman, Gilbert. (1999). "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99: 315–331.

———. (2000). "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society* 100: 223–226.

Hollinger, R., and J. Clark. (1983). *Theft by Employees*. Lexington, KY: Lexington Books.

Hursthouse, Rosalind. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.

Jones, J., K. Slora, and M. Boye. (1990). "Theft Reduction through Personnel Selection: A Control Group Design in the Supermarket Industry." *Journal of Business and Psychology* 5: 275–279.

Kamp, J., and P. Brooks. (1991). "Perceived Organizational Climate and Employee Counterproductivity." *Journal of Business and Psychology* 5: 447–458.

Kant, Immanuel. (2002). *Groundwork for the Metaphysics of Morals*. Translated by A. Zweig. Oxford: Oxford University Press.

Kulas, J., J. McInnerney, R. DeMuth, and V. Jadwinski. (2007). "Employee Satisfaction and Theft: Testing Climate Perceptions as a Mediator." *Journal of Psychology* 141: 389–402.

Miller, Christian. (2013). *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.

———. (2014). *Character and Moral Psychology*. Oxford: Oxford University Press.

Moriarty, T. (1975). "Crime, Commitment, and the Responsive Bystander: Two Field Experiments." *Journal of Personality and Social Psychology* 31: 370–376.

Schwartz, S. (1973). "Normative Explanations of Helping Behavior: A Critique, Proposal, and Empirical Test." *Journal of Experimental Social Psychology* 9: 349–364.

———. (1977). "Normative Influences on Altruism." In *Advances in Experimental Social Psychology*, vol. 10, ed. L. Berkowitz, 221–279. New York: Academic Press.

Slora, K. (1989). "An Empirical Approach to Determining Employee Deviance Base Rates." *Journal of Business and Psychology* 4: 199–219.

Valdesolo, P., and D. DeSteno. (2007). "Moral Hypocrisy: Social Groups and the Flexibility of Virtue." *Psychological Science* 18: 689–690.

———. (2008). "The Duality of Virtue: Deconstructing the Moral Hypocrite." *Journal of Experimental Social Psychology* 44: 1334–1338.

Vranas, Peter. (2005). "The Indeterminacy Paradox: Character Evaluations and Human Psychology." *Noûs* 39: 1–42.

Watson, G. and F. Sheikh. (2008). "Normative Self-Interest or Moral Hypocrisy? The Importance of Context." *Journal of Business Ethics* 77: 259–269.

White, G. (1972). "Immediate and Deferred Effects of Model Observation and Guided and Unguided Rehearsal on Donating and Stealing." *Journal of Personality and Social Psychology* 21: 139–148.

Wicklund, R. (1975). "Objective Self-Awareness." In *Advances in Experimental Social Psychology*, vol. 8, ed. L. Berkowitz, 233–275. New York: Academic Press.

Wimbush, J., and D. Dalton. (1997). "Base Rate for Employee Theft: Convergence of Multiple Methods." *Journal of Applied Psychology* 82: 756–763.

8 }

Emotion and Character

Charles Starkey

I. Introduction

How is emotion related to character? In the domain of philosophy, the study of character typically focuses on character traits—that is, dispositions that are generally taken to be indicative of what kind of person someone is. This conception of character is descriptive rather than evaluative: character traits may be good or bad, desirable or undesirable, but in either case they are a particular kind of feature of a person. One might expect that philosophy offers a robust discussion of the relation between emotion and character traits. However, this is not the case. Though philosophical interest in both emotion and character has exploded in recent years, there has not been a sustained analysis of their relation.

Part of the reason for this is that philosophers have focused on virtue rather than character, leaving unattended the broader relations between emotion and character.[1] Another reason is that the resurgence of virtue ethics that began in the late 1950s has been primarily influenced by the ancient Greek philosopher Aristotle, and philosophers have largely taken Aristotle's comments on the role of emotion in the virtuous person as a starting point, rather than as a position to be scrutinized and defended against various alternatives. The result has typically been a focus on developing a conception of emotion that is consistent with Aristotle's claim that a virtuous person has

[1] Between 1950 and 2000 there were nearly five times more publications on virtue than on character, based on a keyword search in *Philosophers Index*, the main inventory of philosophical work on character. Virtues are commonly considered a species of character trait—namely, the class of good or optimal character traits. For reasons I can't elaborate on here, I believe that virtues are not character traits (i.e., they are not conceptually related) but that character traits have a very important contingent relation to virtues (Starkey 2014). So in either event, understanding character traits and their foundation is an important task of moral psychology.

appropriate emotions, rather than on considering, more fundamentally, *why* this might be the case.[2]

The relation between emotion and character traits is something that we can't take for granted. Even if we believe that emotions bear on character traits, this can take many possible forms. Emotions might be a necessary component of what character traits are. However, this doesn't seem to be the case because emotion is not part of the concept of character in the way that it is, for instance, part of the concept of joy: we can imagine character traits without emotion in a way that we can't imagine joy without emotion. Another possibility is that emotions are relevant to character because they reflect character; that is, they are an indicator of someone's character traits. A person's arrogance may be revealed by his or her anger at reasonable assertions made by colleagues. But if this is the case, then emotions serve to provide other people with information about the person's character, which has nothing in itself to do with that person's character. A third possibility is that emotions underlie the possession of character traits because they produce or support such traits.

In what follows I'll propose something in line with this last option by arguing that emotions have two central roles to play in the possession of character traits. On a *proximal* level, our emotional profile—the range of emotions that we have and the occasions in which they arise—can be a determinant of character by regulating the content of our perception because such perception affects our character traits.[3] The content of such perception includes what we perceive and the significance that such perceptions have to us. On a *fundamental* level, emotions underlie the persistence of character traits by preventing "axiological entropy"—the diminishment over time of the sense of the importance of values on which character traits are based. Without emotions the significance of values related to those emotions would fade and the character traits that involve those values would diminish.[4]

The chapter will address each of these two roles in turn. First, it will identify four ways in which emotions affect our perception and thus our cognition of the things, people, and events that we encounter. These four include

---

[2] See, for instance, Rosalind Hursthouse's (1999) discussion of virtue and the emotions in *On Virtue Ethics*. Other recent work proposes that emotions are relevant to virtue because they skew attention (Goldie, 2004; Sherman, 1989). But this is insufficient because empirical research demonstrates that emotions can have the opposite effect—that they broaden attention. This underscores the point that the work on emotion and character should be informed by and consistent with the relevant empirical psychology, as has been the case in the recent debate over situationism.

[3] By *perception* I mean cognitively "thick" perception, where we come to a robust representation or interpretive conception of the object of perception, not minimal sensory perception. We see a particular pattern of lines and shades on a wooded path in front of us as a snake or a stick, not as a pattern of lines, etc. And we do so immediately rather than as a result of an inference. See Robert C. Roberts's (2003) concept of construal for a related account.

[4] I won't address the demonstrated relation between emotion and motivation here. The interaction between the two is complex but uncontroversial. See Cunningham, Steinberg, and Grev (1980).

attention, perceptibility, information processing, and import. It will then outline the nature and function of axiological entropy, providing evidence that emotions prevent such axiological entropy and, in doing so, serve to sustain character traits. Following this, it will consider several possible objections to these claims about emotion and character.

II. Emotion and Perception

Emotions are among the determinants of character traits on account of the effects of emotion on perception. Emotional responses underlie patterns of perception and, as a consequence, cognition and action that are features of character traits. I will identify four specific ways in which emotions affect perception and consider the impact of each on character traits.

First, emotions affect the scope of our attention. There is a long tradition of psychological work on negative emotions and attention demonstrating that in states of emotional arousal the field of awareness is narrowed or funneled (Easterbrook, 1959). When in an emotional state, participants in one study were less able to report lights that briefly appeared in the peripheral parts of a display than when in an unemotional state, indicating that the focus of attention is broader in unemotional states. Moreover, when participants were in a state of strong emotional arousal, there was no recognition of the peripheral lights. This indicates that the object of perception garners more attention in certain emotional states. More recent work provides evidence that emotionally significant stimuli are the subject of more perceptual processing, and that they are thus identified more easily than unemotional stimuli (Zeelenberg, Wagenmakers, & Rotteveel, 2006). In addition, there is evidence that emotions facilitate attention by increasing "contrast sensitivity," the ability of visual systems to distinguish an object from its background.[5] This increase in contrast sensitivity is independent of attentional focus on the object, but it can make such focus possible. Thus, there are several ways in which emotions affect the scope of attention and the ease with which objects are attended to. This indicates that if we are emotionally aroused by certain types of character-relevant actions or action possibilities, then we will be more likely to focus on them rather than on other parts of the mass of sensory information that we take in and that this focus will be more intense in certain emotional states. So, for instance, we will be more likely to

[5] In one study, participants were asked to judge the orientation of an object as the contrast between it and the background varied. The level of contrast needed to correctly report the orientation of the object was lower in the condition when participants were shown a fearful image for a fraction of a second prior to judging the orientation than when shown an emotionally neutral image (Phelps, Ling, & Carrasco, 2006).

pay attention to the negative effects on other people of a possible course of action if we are emotionally aroused by such a prospect.[6]

Second, emotions affect perceptibility, which includes how a given object of perception appears to us and whether we perceive an object at all. One recent study shows that when people see an ambiguous figure that has a positive association with one possible way of perceiving the figure and a negative association with the other way, they will perceive it in line with the positive association (Balcetis & Dunning, 2006). In this study, the participants were shown the figure very quickly so that they did not have time to recognize it as an ambiguous figure, and thus were not consciously choosing which way to perceive the object. This indicates that emotional states can have a significant effect on, quite literally, what we see when we attend to an object. Regarding character, it provides evidence that, for instance, a sympathetic person will be more likely to perceive an ambiguous situation in one way, and a contemptful person will be more likely to see it another way, such as seeing someone as a victim of circumstances versus seeing them as deserving of their station in life.

In addition, whether an object is perceived can depend on emotion relevance. This is demonstrated by the effect on "attentional blink," the phenomenon where people often fail to detect the appearance of an object when it is presented a fraction of a second after another object in the same space. The visibility of the second object is increased when that object is emotionally relevant, such as a fearful face, and the visibility of the second object is decreased if the first object is emotionally relevant (Anderson & Phelps, 2001; Schwabe & Wolf, 2010). Thus emotion relevance enables us to perceive objects in a given area of attention that we might otherwise not perceive. An implication of this is that in a complex situation with a succession of stimuli, someone who is emotionally aroused by the distress of another person will be more likely to spot someone in distress than someone who is not emotionally aroused. This can promote a character trait of compassion by making compassionate behavior more likely. This is distinct from the attentional aspect discussed earlier because here, emotion affects whether the object is seen rather than how strong the attention to the object is.

Third, emotions alter the way information about our environment is taken in and processed. The psychological function of emotion has been characterized in terms of information processing as an "interrupt system" that sets aside programs that are currently running and shifts information processing resources to important real-time needs. The capacity for such shifting allows the management of multiple goals and their hierarchy (Simon, 1967). On this

[6] There is a growing consensus that some positive emotional states broaden the scope of attention (Fredrickson, 1998; Gable & Harmon-Jones, 2010). This is relevant to character because it can allow a person to notice more of what is happening in his or her environment, so that a potentially generous or courageous person is more likely to notice when someone is in need and thus more likely to be able to act in line with those traits.

account, emotions can interrupt an operative goal program and replace it with a new goal program. Recent research supports something like this view: initial states of emotional arousal can direct further processing of information that the arousal is associated with and thus increase processing of information about the object of the emotion (Le Doux, 1996; Zeelenberg et al., 2006). One example of this is the ability of an emotional response to voice tone to affect cognitive processes before what is said is fully comprehended (Kitayama & Howard, 1994). The enhanced processing of emotionally relevant stimuli includes both the creation and recall of memories relevant to the stimuli. Moreover, the memories created in this way are often more vivid and accurate than unemotional memories, and thus can have a greater bearing on subsequent information processing (Reisberg & Hertel, 2005). Emotional arousal at perceiving a risky but beneficial course of action can increase the direction of cognitive resources to the situation, which is likely to encourage the gathering of more information and facilitate problem-solving directed at successfully acting in the face of the risks. Similarly, an emotion directed at the goal of an act as opposed to the risk can increase the likelihood of courageous behavior.

Fourth, emotions affect the import of the object of the emotion. We continually encounter objects and events, but they do not have the same import to us. "Import" refers to the experienced significance of an object of perception, which includes, in part, the interpretive meaning of the object and inferences about its relation to our goals, desires, and so on. One influential conception of emotion that supports this is the "affect is information" model, according to which emotions provide information about the relation of the object of the emotion to our goals and desires (Clore & Huntsinger, 2007).[7] So our discomfort at seeing someone else mistreated can provide us with information about our attitude toward fair treatment. On this account the information is, in part, embodied: the affective changes in the emotion state are associated with its object and so the emotion becomes part of the information presented about the object. The fear of a snake on the walking path becomes part of the representation of the snake. Here there is not a simple divide between the emotion and the cognition or understanding of the object, and the emotional representation is forceful and provides the object with a significance that it would lack otherwise. For this reason, it has been argued that emotions cut across the distinction between the content of a statement (the information that a statement conveys) and the force of the statement (Gunther, 2004). This is independent of the effects of emotion on attention, information processing, and the like, discussed earlier, as we can pay attention to something, such as a brightly colored object or other novel stimulus, without it having any significance for us. Returning to the example of fair treatment, the pairing of the emotional response with the

[7] For a similar view see Elgin (2008).

recognition of this event can also give it more gravity than it would have otherwise and can thus encourage us to respond to the mistreatment.

Another aspect of this import is the role that emotions have in conveying the significance of action possibilities. Zadra and Clore (2011) argue that "emotions provide embodied information about the costs and benefits of anticipated action, information that can be used automatically and immediately, circumventing the need for cogitating on the possible consequences of potential actions" (p. 676). Emotions thus provide information about not only external objects of perception but also the mental conception of possible courses of action. Feeling fear as we walk toward a cliff provides immediate information about the consequences of continuing to walk forward. Someone's discomfort at the prospect of dodging the truth in a conversation can provide information about the cost of doing so (in terms of compromising that person's values), and in this way provide immediate information about it even before he or she explicitly thinks about the consequences. Emotional responses can work in the other direction as well: the discomfort can be paired with the unpleasant consequences of being truthful. In this latter case the emotion could increase the likelihood of behaving dishonestly. In either event, the emotional responses affect the associated character trait of honesty or dishonesty.

In summary, emotions affect where attention is directed, what is seen when attention is directed, the amount of information gathered about the object of emotion, and the experiential significance that the object has for us. This has implications for character traits because collectively, emotions significantly alter our perception of our environment as we move through the world and, accordingly, play a significant role in determining perceptual dispositions that underlie character-relevant cognition and behavior.

The principal aim of this discussion has been to substantiate the important links between emotion and perception and, thus, character traits. However, this has clear implications for morality and, most pertinent to the topic here, morally relevant character traits. Someone who is emotionally aroused by the opportunity for personal advancement in relation to others is more likely to perceive a situation as one involving an opportunity for such advancement, is more likely to attend to and gain more information about the relevant situation, and will see the opportunity as more important than if there were no such arousal; such a person will be more likely to take the opportunity for the sake of personal advancement and thus to perform an act with the act-character of competitiveness. Likewise, somebody who is emotionally aroused by helping others will have a different perception of an opportunity to perform some community service than someone who is emotionally aroused by the visibility that such service will bring to them, and the act-character of each will be different as a result of the different perceptual and motivational sets.

Possessing a certain sort of character, then, involves possessing, among other things, a certain emotional repertoire. Likewise, aspiring to a certain

sort of character involves developing an emotional repertoire that promotes the proper sorts of perceptions, and working against responses—or their absence—that prevent such perceptions. We should feel a certain amount of anger toward injustice to facilitate the recognition of instances of injustice and to have the proper sense of their significance, and also feel a certain amount of sympathy toward those suffering from injustice. It isn't a matter of simply having or not having the emotional response, because excessive emotion can create false perceptions or excessive sensitivities to conditions and thus a tendency to perceive and cognize situations in line with a certain emotion, even when the perception and cognition are not justified. Someone with an excessive emotional response to disapproval may see situations as instances of disapproval when in fact they are not.[8] Thus, the goal of emotional development is appropriate emotional balance.

III. Emotion, Value, and Axiological Entropy

In addition to affecting character-relevant perception, emotions play a role in fundamentally sustaining character traits. This aspect of the relation between emotion and character is based on a reasonably uncontroversial assertion that I will only briefly outline here—namely, that character traits are dispositions to a range of attitudes and actions that instantiate certain values. Our attitudes and choices are typically determined, when there is a plurality of possibilities, by the level of importance we attach to the various choices we perceive, that is, our values, and the decisions we make between competing interests are indicative of our character traits. Choosing to help a friend in need rather than enjoying one's leisure time is indicative of one sort of trait. Character traits are thus dependent on our values when we evaluate and choose. A generous person is someone who in part is disposed to a range of actions that instantiate the value of giving more than is necessary or expected to others without the expectation of receiving something in return; and an honest person is someone who in part is disposed to a range of actions that instantiate the value of truth. Traits of character are thus different from other traits in that traits of character necessarily involve our values in a way that other traits do not.

AXIOLOGICAL ENTROPY

Here, I argue that emotions are necessary to the sustenance of character traits because emotions prevent "axiological entropy"—the diminution over time

[8] See Amélie Rorty's (1978) discussion of magnetizing dispositions.

of the tacit sense of the importance of values on which character traits are based. Without the ongoing reinforcement of emotions, we would not be able to preserve the sense of the importance of values. Once again, the decisions we make between competing interests are indicative of our character traits. Such decisions, in turn, are determined by our values. Honesty involves in large part the value of truth outweighing the value of personal benefit: an honest person values truth more than personal benefit. In the ongoing absence of relevant emotions, a value such as truth would cease to be important to a person, and would thus be unable to drive the person's perceptions, choices, and actions. They would be less likely to choose and act for the sake of truth, and the corresponding character trait of honesty would be diminished and, over time, extinguished.

There are several lines of support for the phenomenon of axiological entropy and its effect on character traits. One stems from reflection on the relation between emotion and belief. In a discussion of the Stoic view of emotion, Martha Nussbaum refers to the "freshness" of propositions (what I will call "beliefs," for sake of simplicity) that are given a significance by an emotional state. Her central example is grief at the death of a loved one. The fact of the death—the belief that the loved one is dead—is very significant and at the forefront of our mind when we feel grief. The emotion subsides as time passes and with it the level of significance of the death relative to other things; that is, to other beliefs that are in our mind. She explains this diminution of emotional force as the fading of the freshness of the belief. Over time, the belief regarding the death loses "its extreme sharpness, its intrusive cutting edge," and in relation to other beliefs, "[i]t does not assault the other[s] . . . it sits alongside of them" (Nussbaum, 1994, p. 382). The belief no longer dominates other beliefs that are in our mind. The same can be said of other emotionally laden thoughts. The pride we feel at winning an award places the fact of the award in the forefront of our mind, but over time, the fact of the accomplishment diminishes in significance, and it no longer occupies the same place in our mind. The larger point is that we not only have various beliefs, but these beliefs have a certain level of significance to us and this level is variable.

Nussbaum's discussion provides a vehicle for understanding axiological entropy. There is reason to believe that the emotional highlighting and diminution captured by the idea of freshness and fading can occur with the values on which our character traits are based. The particular occurrence of an emotion can enable the specific object of the emotion to have an adequate significance to us, but it can also support the significance of broader values. For instance, feeling pride about a particular accomplishment may give us an adequate sense of the importance of our accomplishment, but the emotion can also more generally support our sense of the importance of accomplishments, so that we are likely to have a similar sense of the value of accomplishment in

the future. Conversely, their importance relative to the many other concepts and values that we deal with will lessen in time—they will lose their freshness. Were we to go through years without feeling pride or satisfaction in our accomplishments, we would not attach this significance to possible accomplishments. In this way, the sense of the significance of particular types of values, in other words, their importance to us, will diminish over time.[9]

CONDITIONING

This characterization of axiological entropy is supported by conditioning theory. Both the pairing of emotional responses and values and the concept of flux in the significance of values corresponds with the principles of classical and operant conditioning, particularly the concept of the extinction of learned responses. I'll review these two types of conditioning before discussing extinction and their relation to axiological entropy.

The attraction (and aversion) to values can involve classical conditioning—the pairing of a neutral stimulus with an unconditioned response to transform the neutral stimulus into a conditioned stimulus. A noted example is that of Albert, a boy who at the age of nine months was exposed to a white rat. He initially displayed no fear of the rat, but in repeated exposures the presence of the rat was paired with a metal pipe hit with a hammer, a jarring sound that scared the boy and made him cry. After repeated pairings, Albert would cry when exposed to the rat without the noise. The unconditioned stimulus (the harsh noise) brought on the unconditioned response (fear), and as a result of conditioning, the rat was transformed from a neutral to a conditioned stimulus, which was accompanied by the conditioned response of fear of the white rat (Watson & Rayner, 1920). Discussions of studies like this focus on the production of an emotional reaction, but it is also the case that the rat became an aversive stimulus, one that provided motivation for avoidance. In addition, such studies involve conditioning with specific objects, but, as I'll discuss later, after reviewing operant conditioning, this conditioning can include broader properties, including general values.

Such aversion and attraction can also result from operant conditioning. Operant conditioning involves voluntary behavior that is rewarded with incentives or punished with disincentives (Skinner 1948). Behaviors or behavioral dispositions are developed or diminished through such

[9] Nussbaum notes that the ancient philosopher Chrysippus holds that in order to form passions, judgments must be "fresh," and that "the point of this seems to be to allow for certain sorts of affective distancing, especially over time, compatibly with the retention of the same belief or judgment." This is similar to my claim that concepts and values lose their force, except that I hold that a belief or judgment with a different level of significance is not entirely the same belief. See Nussbaum (1994), p. 381.

reinforcement or punishment. Here, rewards and punishments are understood broadly; so someone can be punished for cutting in line with facial expressions, gestures, and verbal condemnation rather than through a formal penalty, and can be rewarded in similar ways. Operant conditioning comes out of the behaviorist tradition, but in considering its role here, we need not limit it to the conditioning of observable behavior. People display "reactive attitudes" of gratitude and anger in response to the perceived affection and good will or contempt and malevolence of others (Strawson, 1974). In this way, reactive attitudes can reinforce or deter certain attitudes and intentions. Praising someone's telling the truth for the sake of being honest can reinforce that intention, as can the positive feelings within that person that can result from their thinking of their intentions as good intentions that are approved by others. The same can be the case with negative feedback directed at bad intentions. Thus, through this sort of conditioning people can be encouraged to act not only in certain ways but with certain motives and intentions. This is relevant to character because acts of a certain character involve specific intentions: giving more than is necessary or expected to someone for the sake of that person's good is generous, but giving more than is necessary or expected because of anticipated benefits to oneself is not generous.

The features of classical and operant conditioning shed light on axiological entropy because conditioning creates attractions and aversions to not only the specific object involved in the conditioning process, but also, through *stimulus generalization*, to more general properties possessed by that object, and these properties can include values. Stimulus generalization supports the relation between Nussbaum's concept of freshness and general values proposed earlier. This phenomenon was observed in the Albert study. After the fear conditioning toward the white rat, Albert became fearful of a variety of other white objects as well as the rat. An aversion had been created toward the broader property of "white." Stimulus generalization can occur in cases where the objects have minimal physical similarity; here the objects share the same stimulus concept, though little in the way of physical features (Martin & Pear, 2010). In this way, two instances of generosity can share the same conceptual features of a generous act, but otherwise have different characteristics, and a positive emotional response to an episode of generosity can inform the evaluation of, not only the specific instance of generosity, but also the broader property of generosity, and thus affect a broader range of decisions where generosity is at stake.

Thus, these two types of conditioning constitute two different mechanisms through which emotional conditioning can affect the level of importance of values. The associated phenomenon of extinction, which is considered next, counters conditioning and provides a theoretical grounding for the operation of axiological entropy.

EXTINCTION

The converse of conditioning is extinction. In a way resembling Nussbaum's description of the fading of the force of beliefs over time, the repeated presence of a conditioned stimulus in the absence of an unconditioned stimulus will decrease the occurrence of the conditioned response. This is a feature of emotional as well as behavioral responses. Affective responses to stimuli diminish over time, as studies on the "fading effect" have shown (Holmes, 1970; Walker, Vogl, & Thompson, 1997). This can also include our sense of the importance of general concepts or values on which our character traits are based. Over time, their importance relative to the many other concepts and values that we deal with will lessen—they will lose the freshness that Nussbaum refers to. Thus, the import and value of particular classes of goods can be lost.

The implication of this is that, in an environment of competing possible observations, situations that involve values that have diminished will not only be less noticeable; they will not have the same weight when standing against competing values that they had before the diminution. If the value of accomplishment diminishes, we will not value particular accomplishments as much and will be less likely to notice and care about situations in which accomplishment is at issue. This is not to deny that judgments could be made in such cases. However, we would not be as likely to recognize an accomplishment were the value of accomplishment to lose its freshness, and such a value or the judgments that relate to it would have little significance. Similarly, in the absence of reinforcement, the value of honesty would diminish and eventually cease to be of particular significance to a person, and would thus be unable to drive the person's perceptions, choices, and actions. He or she would be less likely to perceive instances of honesty or dishonesty as such, and would be less likely to choose and act for the sake of truth. The corresponding character trait of honesty would be diminished or extinguished. This diminution of the sense of the importance of general values, if unchecked, will degrade both our ability to see various things that should be made salient because they involve those values, and the ability of our practical reason to function properly.[10] In this way, without our periodically having the relevant emotions, the significance or weights attached to particular values would fade and the character traits that involve those values would cease to be.

Support for this conception of the relation between emotion and value comes from research on post-traumatic stress disorder (PTSD). One manifestation of PTSD is the phenomenon of emotional numbing, including feelings of detachment and a reduced range of affect (Shay, 1994). Such emotional numbing is associated with callousness and antisocial behavior (Kerig &

[10] Ronald de Sousa discusses the determination of the significance of various facts in practical reasoning and its relation to the "frame problem" in artificial intelligence. See de Sousa (1987).

Becker, 2010; Allwood, Bell, & Horan, 2011). This broad emotional numbing is consistent with the display of anger, and thus some emotional responses may be numbed, while others are still operative; but the numbing that exists is closely associated with maladaptive behavior. This indicates that without the support of emotions, the values that underlie character traits lose their force.[11]

COUNTERING AXIOLOGICAL ENTROPY

Of course, in the proper amount, this diminution of the importance of particular objects is a good thing. If the pride of winning a grade-school spelling contest filled our mind and caused us to gloat in the recollection of the accomplishment a decade later, we would be distracted from our present work and from many other things that are, at this later place and time, much more important. Similarly, mourning over the death a friend should subside over time to enable us to deal with our other concerns, though the death is still a tragic loss.

However, if unchecked, this diminution will, once again, affect the importance to us of things that should remain valuable, particularly certain types of things, including the values associated with character traits. Though much of the discussion of psychological conditioning centers on *creating* associative links and responses, the same principles apply to *maintaining* such responses, that is, keeping existing responses from fading. What prevents this axiological entropy associated with extinction from occurring is the successive emotional pairing with a value—the having of emotions that are directed at objects that involve these values and serve to maintain or strengthen our sense of their importance. Here, the having of relevant emotions from time to time reinforces or refreshes general values through stimulus generalization. To return to examples cited earlier, associating negative emotional arousal with an instance of harm gives the broader concept of harm a greater weight than it would have otherwise, and it informs our evaluation of situations where harm or the risk of harm is at issue. In the same way, the association of negative emotion with episodes of dishonesty gives the value of truth a particular significance, including in decisions where truth is at stake.

This emotional pairing can take several forms. An emotional reaction to a perceived value can produce the pairing of emotion and value in a way that reflects classical conditioning. We may see someone behaving callously to another person, which arouses indignance on our part directed at the

[11] There is reason to believe that this fading of the level of importance of values can take place at a larger cultural level as well as an individual level. The decline of honor culture in the West (Bowman, 2007; Wyatt-Brown, 1982) illustrates this. Over the past century honor has become less important in Western culture, and this can be explained as the decreasing emotional resonance that issues involving traditional concepts of honor have. Thus axiological entropy can affect larger cultural values.

callous act, reinforces within us the sense of the significance of callousness, and sustains the aversion to callous acts. In addition, the expression of emotion by people can elicit similar feelings in other people directed at the same object—what Bernard Rimé describes as "emotional communion." He notes that "[t]he shared episode and the expression of associated emotions by a member of the social group have the power to vividly elicit analogous feelings in people around them, so that a reciprocal stimulation of emotion follows" (Rimé, 2013, p. 95).

Observing the reactive attitudes of others toward what we do functions in a similar way. This is different from emotional communion because here, the reinforcing emotion within us is not the same as the emotional reaction of other people. For instance, negative reactive attitudes (anger and resentment) in response to a rude act on our part can elicit shame in us and, in line with classical conditioning, produce an emotional association with the act, so that the thought of that type of act itself is aversive. And in line with operant conditioning, this can also produce an association with the negative *consequence* of the act that deters us from acting in that way. For instance, when we behave rudely and feel shame after observing the anger of other people toward the behavior, the shame that is induced marks the consequences of rudeness (the condemnation of others) with a negative association. Positive reactions to behavior, such as gratitude, operate in the same way, but produce positive associations with the type of act and its consequences.

Another device is the experience of art and literature, where emotional reactions are encouraged through contextual devices—music, lighting, the appearance of the people involved, and so on. All of these serve to prompt and support particular emotional responses directed at objects with particular values. In the film *The Sweet Smell of Success*, the dark, atmospheric camerawork and Burt Lancaster's icy vocal delivery as the semifictive columnist J. J. Hunsecker work together to heighten our chilly distain at the megalomania that is at the core of Hunsecker's character. More generally, art and literature are valuable because of their ability to induce emotions and pair them with values.[12]

Support for the role of emotions in preventing axiological entropy comes from studies of clinical psychopaths showing that they display both a lack of emotion to adverse events and diminished ability to empathize (Blair, Colledge, Murray, & Mitchell, 2001; Blair, Jones, & Clark, 1997; Kiehl et al., 2001). Along with this emotional deficit, they lack, as Patricia Greenspan (2003) notes, "the normal measure of receptivity to moral reasons—the capacity to appreciate their force as moral reasons, which involves the capacity for a certain kind of immediate emotional response"(p. 420). Note that

[12] See, for instance, Ronald de Sousa's (1987) discussion of the reinforcement of "paradigm scenarios" provided by art and literature.

what appears to be present in such psychopathology is the diminished or missing *capacity* for certain emotional responses, not merely the absence of emotional response in many specific situations where such a response would be appropriate. Greenspan's reference to appreciating the force of moral reasons is telling: though her focus is on deficits in the emotional response to moral norms, the point applies to character traits as well. It follows from this model that people with an impaired capacity for emotional response will have an impaired moral character because their emotions do not function normally to prevent axiological entropy. The research on psychopaths is not only consistent with this but provides support for the model in their diminished appreciation of moral concepts, which appears to stem from the diminished emotional capacity, rather than vice versa.[13] Moreover, impulsivity, a hallmark feature of psychopathy, can be explained by the weak set of values that would keep impulses in check.

Thus, emotional association with values counters axiological entropy and maintains or enhances the importance these values have for us. Such values include the general values that character traits are based on. A consequence of this is that the maintenance of beneficial character traits takes place at both an individual and a cultural level. Part of character improvement involves recognizing and fostering cultural forces that support the values that underlie character traits and the correct level of importance of these values. As the foregoing discussion suggests, this can involve eliciting appropriate emotions in ourselves and others and the social sharing of emotions in the form of emotional communion. These are not sufficient for the possession of beneficial character traits, as it also requires the ability to discern when and how these values apply to particular situations. This ability can be provided in part by the role of emotions in regulating our perception that has been discussed. It also requires practical wisdom in adjudicating between different values, such as when compassion and justice conflict. But though the proper emotional repertoire is not sufficient, it is necessary for possession of these traits.

IV. Objections

One possible objection to the ideas presented here is that emotions reflect perceptual states rather than cause them. So, emotions do not determine character traits, but rather character traits and their associated perceptual states

[13] People with localized damage to the brain regions associated with the emotions tend to not have the same response to moral scenarios that involve what are typically emotion-inducing events, though their cognitive functioning appears to be normal. See Koenigs et al. (2007). In addition, the perceptual effects of emotion discussed earlier result from the emotional state and are not prior to the emotional state, and these can affect the significance of such scenarios for the perceiver.

determine emotion. For instance, our disposition to recognize instances of injustice might make us angry at such instances, rather than our anger helping us recognize the injustice. In response, the first claim is correct but the second one is too strong. Emotions do certainly reflect perceptual states because recognizing to some degree an emotionally arousing object is a precursor to having an emotion. But perception is an ongoing process and, as indicated in the literature cited earlier, emotions change the nature of the perception by altering various aspects of the processing of information and the force of the representation. The relation between emotion and perception is dynamic, so that the perception of properties can activate emotion systems, but emotion systems in turn modify what is perceived (Le Doux, 1996).

Another objection is that responsiveness to basic values is innate, so emotional reinforcement is not needed for such values to be significant and is thus not a requirement for the possession of character traits. One study concludes that three-month-old infants show an aversion to antisocial actors (Hamlin, Wynn, & Bloom, 2010). Another study concludes that six- to ten-month-olds have a strong preference for prosocial (nice) actors over antisocial (mean) actors (Hamlin, Wynn, & Bloom, 2007). Other research provides evidence that children develop a sense of fairness in distribution of resources between twelve and fifteen months (Sommerville, Schmidt, Yun, & Burns, 2013). However, these findings are consistent with emotions playing the fundamental role outlined above. The first two studies are consistent with rudimentary emotion systems marking prosocial and antisocial behavior, and thus consistent with the capacity for emotional response being required for the preference for prosocial actors. Similarly, the findings of the last study are consistent with the necessity of emotional responses to care about fairness in distribution. This possibility is underscored by Paul Bloom (2011), who notes, "What we see in these emerging emotions [in infants] is the development of mattering" (p. 78).

A third objection is that because we can obviously *act* when we are not in an emotional state, we can have character traits without emotion. One can help a friend, arrive at a meeting on time, and perform many other tasks indicative of a certain character without being emotionally aroused, so emotions are not necessary for possessing character traits. This objection is unsuccessful because we perceive and understand the objects and situations around us in terms of various goods and interests, and our perception and motivation is driven by these evaluations. Most long- and short-range planning is based on values, not on simple pains and pleasures. Hunger by itself is a pain, but it is not evaluative; it is not concern or frustration about our hunger. Thus, action that is prompted by hunger or other simple pleasures or pains does not instantiate a character trait: character traits involve acts based on values and broad patterns of perception, cognition, and action that are dependent on emotions for the reasons given earlier. Moreover, our ability to act without emotion can be explained by conditioning, which can affect evaluations and choices even

when there is no occurrent emotional state.[14] In addition, even if emotions were not necessary for some occasional character-relevant acts, they would still be necessary for character *traits* because such traits are based on ongoing dispositions that would be unlikely to exist in the total absence of emotions.

A more substantial criticism of the proposal that emotions play an important role in causing or sustaining character traits is that character involves rational will power—resisting the force of emotions that can compromise our ability to reason or our resolve to do the right thing. G. H. von Wright (1963), for example, believes that this is a mark of the virtues, claiming that "[i]n the case of every specific virtue there is some specific passion which the man of that virtue has learned to master" and that "the role of a virtue, to put it briefly, is to counteract, eliminate, rule out the obscuring effects which emotion may have on our practical judgment" (p. 147). Von Wright does note that it does not seem "unplausible" that emotions may at times serve the good of people, but does not consider this possibility further, nor does he seem inclined to do so: he refers to the issue of this possibility as "bewildering" (p. 147).

The attitude that von Wright exemplifies has its point: reason and emotion often conflict, and in many of these cases *performing* the proper course of action involves ignoring the promptings of our emotions and instead following the dictates of our reason. We may be tempted to flee in the face of danger, though we clearly know that we should face the danger because of the consequences of not doing so. In addition, recognizing the proper course of action and acting accordingly can involve suppressing our emotions or counteracting their "obscuring effect." Extreme anxiety, anger, or even pride can disable our practical reason and make us unable to decide or cause us to lose sight of what we should do. In cases such as these, there is surely a virtue in remaining calm and counteracting the effects of emotions as best we can.

Though von Wright's focus is on virtue, this idea can be extended to character traits by claiming that character traits should be the product of rational reflection and resolve rather than emotional disposition. However, such pure rational resolve is both unrealistic and, ultimately, undesirable. As argued here, emotions play a key role in perception and in maintaining our sense of the importance of various values and their representation as something that we are attracted or averse to. Since emotions are determinative of what we see, take note of, and regard as important, emotions play a part in the formation of rational resolve itself. Thus, a strict rationalist requirement violates the principle of *minimal psychological realism*. According to this principle, the demands of any moral theory should consist of things that humans are at least

[14] One way of explaining this is the somatic marker hypothesis put forward by Antonio Damasio. According to this hypothesis, subtle bodily responses to stimuli and the brain's representation of such responses even in the absence of such responses serve to bias deliberation in line with the responses (Damasio, 1998).

capable of, even if they are difficult—things that are "possible for creatures like us" (Flanagan, 1991, p. 32). This principle applies to a conception of character traits as well: what these traits are or ought to be should be consistent with the psychological constitution of humans.

Moreover, we should not let the negative effects of emotions mask their positive role. We should not, of course, always act according to the promptings of our emotions; but, on the other hand, emotions can direct attention and encourage action in ways that are beneficial and that reason alone is incapable of. Even in cases where we do not, and should not, act according to occurrent emotional prompts, our broader emotional repertoire can be operative in steering us toward what is important. In this way, proper emotions and reason have a positive relation. The apparent conflict between reason and emotion may in fact be a sort of conflict between emotions: a conflict between the prompts of an occurrent emotion, on the one hand, and, on the other hand, a sense of what is important, which is a product of a broader, ongoing emotional repertoire.

In sum, none of these objections to the role of emotions in determining and sustaining character traits are compelling.

V. Conclusion

By affecting our perception and maintaining the importance of our various values, emotions play a critical role in the possession of character traits. The consequences of this, as outlined here, is that character development must involve working toward an appropriate emotional repertoire, including proper emotional balance, and supporting character-relevant values on both an individual and cultural level through eliciting emotions in ourselves and others.

Two concluding observations are in order. First, one might take the position that emotions determine the ends of our actions: we decide and act for the sake of what we are emotionally drawn to and against what we are emotionally repelled by.[15] The account proposed here is both narrower and more complex: emotions set parameters on our deliberation and actions in the service of our ends, whatever the source of our ends may be. By influencing both our perception and deliberation, emotions limit what we conceive of doing and what we eventually do. We could want very much to sell our house and strive to sell it, but we answer questions about its condition accurately because of the importance of honesty to us. In this way emotions direct the process of deliberation itself, which is importantly distinct from setting the ultimate ends of deliberation.

[15] See, for instance, de Sousa (2004).

Second, the arguments in this chapter do not imply that emotions determine character traits in any simple way. We may have excessive emotional reactions, which, if taken as infallible indicators of the importance of their object, would deliver a very inaccurate sense of such importance. But emotionally driven perceptions and action tendencies can be moderated by a wider range of experience or reflective consideration, so that we may not follow the promptings of a particular emotion. Just as one may be disposed to be afraid of dogs and, knowing that this is the case, temper one's sense of the actual danger that dogs pose, one may be emotionally aroused by a value, such as loyalty, and place a premium on it, but may also be reflectively aware of this and temper the motivational salience of loyalty with a broader range of knowledge and experience. Beliefs about the significance of loyalty here do not simply correspond with the degree of one's emotional response to situations where loyalty is at issue. More broadly, emotional responses do not rigidly determine the level of importance of values. Nevertheless, emotions are the groundwork for this sense of importance as well as for the perceptual dispositions that underlie character traits.[16]

References

Allwood, M. A., Bell, D. J., & Horan, J. (2011). Posttrauma numbing of fear, detachment, and arousal predict delinquent behaviors in early adolescence. *Journal of Clinical Child & Adolescent Psychology*, *40*(5), 659–667.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*(6835), 305–9.

Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*(4), 612–13.

Blair, R. J., Colledge, E., Murray, L., & Mitchell, D. G. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal and Child Psychology*, *29*(6), 491–7.

Blair, R., Jones, L., & Clark, F. (1997). *The psychopathic individual: A lack of responsiveness to distress cues?*. *Psychophysiology*, *34*(2), 192–6.

Bloom, P. (2011). Moral nativism and moral psychology. In M. Mikulincer, & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 71–89). Washington DC: American Psychological Association.

Bowman, J. (2007). *Honor: A history*. New York: Encounter.

[16] Thanks are owed to the participants of the 2013 Character Project workshop for valuable discussion and feedback on many of the ideas presented here and to Christian Miller for helpful comments on a draft of this chapter. This paper was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Science*, *11*(9), 393.

Cunningham, M. R., Steinberg, J., & Grev, R. (1980). Wanting to and having to help: Separate motivations for positive mood and guilt-induced helping. *Journal of Personality and Social Psychology*, *38*(2), 181–92.

Damasio, A. (1998). The somatic marker hypothesis and the possible functions of the prefrontal cortex. In Roberts, Angela C. Robbins, Trevor W., & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive function* (pp. 36–50). Oxford: Oxford University Press.

de Sousa, R. (1987). *The rationality of emotion*. Cambridge, MA: MIT Press.

de Sousa, R. (2004). Emotions: What I know, what I'd like to think I know, and what I'd like to think. In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 61–74). Oxford: Oxford University Press.

Easterbrook, J. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183–201.

Elgin, C. (2008). Emotion and understanding. In G. Brun, U. Doguoglu, & D. Kuenzle (Eds.), *Epistemology and emotions* (pp. 33–49). Hampshire, Engl.: Ashgate.

Flanagan, O. J. (1991). *Varieties of moral personality: Ethics and psychological realism*. Cambridge, MA: Harvard University Press.

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology*, *2*(3), 300–19.

Gable, P. A., & Harmon-Jones, E. (2010). The blues broaden, but the nasty narrows: Attentional consequences of negative affects low and high in motivational intensity. *Psychological Science*, *21*(2), 211–14.

Goldie, P. (1998). Review of *What Emotions Really Are. British Journal for the Philosophy of Science*, *49*, 642–648.

Goldie, P. (2004). Emotion, reason and virtue. In D. Evans, & P. Cruse (Eds.), *Emotion, evolution, and rationality* (pp. 249–67). Oxford: Oxford University Press.

Greenspan, P. (2003). Responsible psychopaths. *Philosophical Psychology*, *16*(3), 417–29.

Gunther, Y. (2004). The phenomenology and intentionality of emotion. *Philosophical Studies*, *117*(1–2), 43–55.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–9.

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–6.

Holmes, D. S. (1970). Differential change in affective intensity and the forgetting of unpleasant personal experiences. *Journal of Personality and Social Psychology*, *15*(3), 234–5.

Hursthouse, R. (1999). *On virtue ethics*. Oxford, England: Oxford University Press.

Kerig, P. K., & Becker, S. P. (2010). From internalizing to externalizing: Theoretical models of the processes linking PTSD to juvenile delinquency. In S. J. Egan (Ed.), *Posttraumatic stress disorder (PTSD): Causes, symptoms and treatment* (pp. 33–45). Hauppauge, NY: Nova Science Publishers.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry*, *50*(9), 677–84.

Kitayama, S., & Howard, S. (1994). Affective regulation of perception and comprehension: Amplification and semantic priming. In P. M. Neidenthal, & S. Kitayama (Eds.), *The heart's eye: Emotional influences in perception and attention* (pp. 41–65). San Diego, CA: Academic Press.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908–11.

Le Doux, J. (1996). *The emotional brain*. New York: Simon and Schuster.

Martin, G., & Pear, J. (2010). *Behavior modification*. Upper Saddle River, NJ: Pearson.

Nussbaum, M. (1994). The stoics on the extirpation of the passions. In M. Nussbaum (Ed.), *The therapy of desire* (pp. 359–401). Princeton, NJ: Princeton University Press.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science*, *17*(4), 292–9.

Reisberg, D., & Hertel, P. (2005). *Memory and emotion*. New York: Oxford University Press.

Rimé, B. (2013). Can socially sharing emotions change emotions? In D. Hermans, B. Rimé & B. Mesquita (Eds.), *Changing emotions* (pp. 91–5). New York: Psychology Press.

Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge: Cambridge University Press.

Rorty, A. (1978). Explaining emotions. *Journal of Philosophy*, *75*, 139–61.

Schwabe, L., & Wolf, O. T. (2010). Emotional modulation of the attentional blink: Is there an effect of stress? *Emotion*, *10*(2), 283–5.

Shay, J. (1994). *Achilles in vietnam: Combat trauma and the undoing of character*. New York: Athenaeum.

Sherman, N. (1989). *The fabric of character*. Oxford: Clarendon Press.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, *74*(1), 29–39.

Skinner, B. F. (1948). "Superstition" in the pigeon. *Journal of Experimental Psychology*, *38*(2), 168–72.

Sommerville, J. A., Schmidt, M. F. H., Yun, J., & Burns, M. P. (2013). The development of fairness expectations and prosocial behavior in the second year of life. *Infancy*, *18*(1), 40–66.

Starkey, C. (2014). Virtue without character. Unpublished manuscript.

Strawson, P. F. (1974). *Freedom and resentment and other essays*. London: Methuen.

von Wright, G. H. (1963). *The varieties of goodness*. London: Routledge & Kegan Paul.

Walker, W. R., Vogl, R. J., & Thompson, C. P. (1997). Autobiographical memory: Unpleasantness fades faster than pleasantness over time. *Applied Cognitive Psychology*, *11*(5), 399–413.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(3), 1–13.

Wyatt-Brown, B. (1982). *Southern honor: Ethics and behavior in the old south*. New York: Oxford University Press.

Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: The role of affective information. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 676–85.

Zeelenberg, R., Wagenmakers, E., & Rotteveel, M. (2006). The impact of emotion on perception: Bias or enhanced processing? *Psychological Science*, *17*(4), 287–91.

PART } IV

Character and Ethical Theory

9 }

Taking Moral Risks and Becoming Virtuous

Rebecca Stangl

1. Introduction

Almost all classical forms of virtue theory claim that an active life of virtue is the greatest of goods.[1] The Stoics go even further, arguing that virtue is the *only* thing that is genuinely good (Cicero 2001). And while Aristotle, unlike the Stoics, joins common sense in affirming health, wealth, and honor as genuine goods, he nonetheless holds that an active life of virtue is a greater good than these.[2] In this, many contemporary virtue theorists have followed their classical predecessors. They too hold that an active life of virtue is the greatest of goods (Anscombe 1958; McDowell 1979, 1995; Garcia 1987, 1990).

Suppose virtue ethicists are right about this. Moral agents sometimes face what I shall call a situation of "moral risk": a situation in which an agent is capable of achieving some important non-moral good for others, but only by embarking on a course of action that poses a genuine risk to her character—and so poses a genuine risk to her leading an active life of virtue. It might seem that, if an active life of virtue is the greatest of goods, taking such risks is never permissible. But I argue that, on the contrary, taking moral risks is not only compatible with the claim that an active life of virtue is the greatest of goods, but also that, even given that claim, taking such risks is sometimes required.

I begin, in section 2, by clarifying the notion of a "moral risk." In section 3, I show how situations of moral risk raise a *prima facie* problem for the claim that an active life of virtue is the greatest of goods. In sections 4 and 5, I lay out my positive argument that such a claim is not only compatible with taking

[1] Plato, Aristotle, and the Stoics all seem to have held a version of this view. Cf. Annas (1993), especially chapters 1 and 2.

[2] I mention the Stoic view not to endorse it, but only to emphasize how strong the commitment to the value of virtue is in the classical tradition. With Aristotle, I assume both that an active life of virtue is the greatest of goods and that there are nonetheless other genuine goods.

moral risks but may even require it. Finally, in section 6, I respond to two objections to my argument.

2. The Concept of a "Moral Risk"

Let me begin by clarifying the notion of a "moral risk" and distinguishing it from several closely related concepts with which it might easily be confused. A situation of "moral risk" as I shall use the term, is a situation in which an agent is capable of doing significant non-moral good for others, but only by embarking on a course of action that poses a genuine threat to her own virtue, either by weakening one or more of her virtues, causing the loss of one or more of her virtues, or fostering one or more vices. In doing so, it thus so poses a genuine risk to her leading an active life of virtue. A non-moral good, in this context, is simply a good other than virtue or virtuous activity. So health, even though it is a necessary condition of acting virtuously, is still a non-moral good. Similarly, while temperance and acting temperately are moral goods, food, drink, and sex are non-moral goods. This is so even though one can only act temperately if these latter goods are present. Here is an example of a "moral risk":

> Nancy is thinking of going into nursing, because she sincerely wishes to relieve the pain and suffering of others. And were she to go into nursing, she almost certainly would relieve much pain and suffering. But Nancy also rightly fears that the exposure to pain and suffering might make her a less compassionate, and thus less virtuous, person. She might even get to the point where she does the job only for the money. If that were to happen, her ability to live an active life of virtue in the future would be seriously compromised. Should she pursue a career in nursing?[3]

Nancy believes, on reasonable grounds, that she could do genuine (non-moral) good by embarking on a particular course of action. But she also fears, on no less reasonable grounds, that the course of action in question will corrupt her character. The question I am interested in is this: supposing that an active life of virtue is the greatest of goods, is it ever permissible for her to take the moral risk involved? And if it is, when?

This is not simply a species of what Michael Walzer (1973) calls "dirty hands" problems. In a "dirty hands" case, the agent is faced with the prospect of achieving something very morally laudable at the price of doing something else that is extraordinarily morally problematic. (If we torture the terrorist to

[3] This is a version of a case discussed by Hurka (1998, 2001), about which I will have more to say in what follows.

get information about the whereabouts of a bomb, we achieve something morally laudable at the price of what is obviously, and very deeply, morally problematic.) But that is not the case here. Nancy is not considering performing an action which is, of itself, morally problematic. Becoming a nurse is not in any way morally problematic; if motivated in the right way, it may be morally laudable. What makes the choice interesting is not the nature of the act chosen, but the relation in which the act stands to her future character.

In some ways, taking a moral risk is closer to a species of moral luck: whether one's choice to become a nurse is justified might seem to depend, retrospectively, on how the life of the individual turns out. Indeed, the very description of the choice might seem to depend upon how the life of the individual turns out. So suppose Nancy retains her commitment to the relief of suffering, her career allows her to make the maximum contribution to that end, and she grows more rather than less compassionate as a result. In this case, we (and she) may come to judge her entrance into the nursing field as a laudable step in pursuit of that end. Suppose, on the other hand, her character becomes so corrupted that she only does the job for the money, and becomes numb to suffering in general. In this case, we (and she) may come to judge her decision to enter the field of nursing as the first step on her road to corruption—the point at which everything took a turn for the worse.[4]

In this way, a moral risk seems similar to the risk undertaken by the Gauguin of Bernard Williams's (1976) famous example.[5] On Williams's telling, whether Gauguin's decision to abandon his family is justified depends, retrospectively, on whether or not he succeeds in doing genuinely great work. If he does, the decision will come to seem regrettable, yet justifiable—a necessary sacrifice. If he does not—and if he does not *because* he lacks talent, and not, for example, because he is drowned at sea—the decision will come to seem self-indulgent and foolish.

But if there are similarities to the problem of moral luck, the difference is more important. The Gauguin case is a paradigmatic case of moral luck not only because we seem to judge his decision to leave his family as justified or unjustified depending, retrospectively, on whether or not he succeeded in doing genuinely valuable work. [6] It is a paradigmatic case of moral luck because whether he succeeds depends almost entirely on whether he is natively gifted, and that lies outside his control. But a moral risk seems importantly different in this respect. What kind of person I become is no doubt affected by the

---

[4] I thank Tal Brewer for very helpful discussions on this point.

[5] I say "the Gauguin of Williams's example" because Williams does not intend the example to be an accurate piece of Gauguin biography. And it is the case as Williams describes it that I am interested in.

[6] Of course, whether the justification for his action actually depends on the success or failure of the outcome is controversial. For an argument that it does not, see Wallace (2013). Thanks to Christian Miller for this reference.

circumstances in which I find myself—that is why we are rightly worried about taking some moral risks. But only the most pessimistic of us would claim that the kind of person I become depends almost entirely on the circumstances in which I find myself. And hence, the kind of person I become does not depend upon factors that are almost entirely outside my control. So the question before us is not reducible to a question about moral luck.

3. The Problem of "Moral Risk" in Virtue Ethics

The example of Nancy is adapted from Thomas Hurka (Hurka 1998, 2001). Hurka raises the example in order to argue against a very strong version of the claim that virtue is the greatest of goods, a version which he calls the lexical view. According to the lexical view, virtue is infinitely more valuable than any other good, and vice is infinitely worse than any non-moral evil. But if virtue were of infinite value, according to Hurka, then it would be wrong for our aspiring nurse Nancy to take up a medical profession even if the chances of this worsening her own character were *infinitesimal*. But this seems absurd. Hurka thus concludes that virtue is not infinitely greater than all other goods.

Of course, one can hold that virtue is the greatest of goods without holding that it is infinitely greater than any other good. So the problem as Hurka raises it is not a problem for virtue ethicists as such—and Hurka, it should be noted, never claims otherwise.[7] Hurka's argument may well give us reason to reject the Stoic claim that virtue is the only thing that is good, and vice the only thing that is evil. If virtue were the only thing that is good, it would be difficult to see how one might permissibly risk her own virtue in pursuit of any non-moral "good" (or, as the Stoics would have it, a "preferred indifferent").

But Aristotelians, along with most other virtue ethicists, explicitly reject this Stoic claim in any event. Nor does there seem any compelling reason to ascribe the lexical view to them on other grounds. This might not be as clear as it should because the second part of the lexical view—that vice is infinitely worse than any non-moral evil—can be mistaken for a related, but not identical, claim: that some vicious actions are never to be done and can never be justified no matter what non-moral good they can achieve. Aristotle famously claims that theft, murder, and adultery are all actions whose very name connotes depravity, and which can never be right no matter what the circumstances (1999, 1107a8–12).

[7] Hurka does suggest that virtue ethicists are committed to a claim about the priority of virtue whose "simplest reading" is equivalent to the lexical view. But he allows that other interpretations are possible. See Hurka (2001), 240-3.

But one can hold that some vicious actions are never to be done without holding that any amount of vice is worse than any non-moral evil. For one thing, a vice might never eventuate in any action at all. For another, to say that no amount of non-moral good could ever justify some particular set of vicious actions does not imply that no vicious actions can be so justified. Given this, I see no reason to ascribe the lexical view to Aristotelians. They can reject this extreme claim in favor of the more plausible claim that an active life of virtue, while the greatest of goods, is not infinitely greater than all others. So, in what follows, I shall henceforth adopt this weaker claim.

But a case like Nancy's might still seem to present a problem for the virtue ethicist. Even if an active life of virtue is only a finitely greater good than all other goods, it might still seem to follow that Nancy ought not to risk becoming less compassionate in order to secure a healthier life for someone else. After all, the virtue ethicist claims that an active life of virtue is more important than health. So how can it be right to risk the former in pursuit of the latter?

But this seems implausible. Nancy, our aspiring nurse, can do much good in the world. Refusing to do it out of a fear she may become less compassionate seems to reflect a myopic obsession with her own character. Indeed, it might seem to involve the virtue theorist in a kind of practical contradiction: If Nancy's own compassion is more valuable than someone else's health, then it seems she should not risk the former to promote the latter. But isn't the point of being compassionate that we care about other people and not merely ourselves?

In what follows, I argue that, contrary to appearances, the virtue ethicist is not committed to denying that Nancy can ever take the moral risk in question. Taking moral risks is not only compatible with the claim that an active life of virtue is the greatest of goods, but also, even given that claim, taking such risks is sometimes required.

But before proceeding to the argument, two clarifications regarding the claim are in order. First, the claim is that an active life of virtue is the greatest of goods; this is not equivalent to the claim that virtue is the greatest good. To say that an active life of virtue is the greatest good is to say that one's life goes better, and is more excellent, just insofar as one acts from the virtues. It is not to say that virtuous dispositions, understood as enduring character traits, are the greatest goods; nor is it to say that one ought always to maximize the existence of such states. Of course, an active life of virtue can only be lived by someone who is in fact virtuous. So the possession of a virtuous character is necessary to achieve the highest good. But the distinction between the value of virtue and the value of an active life of virtue is nonetheless important, as we will see.

Secondly, to say that something is the highest good is not necessarily to say that it ought to be maximized (or even promoted). Virtue ethicists need not—and typically do not—accept such a tight connection between value and

right action.[8] Like deontological theories, classical forms of virtue ethics recognize a gap between what is of value and what it is right to bring about. To act in accord with the virtue of justice, for example, is not to act in such a way as to maximize just outcomes (even at the price, for example, of committing an injustice oneself). It is, first and foremost, to do what is just. Of course, the virtue ethicist does not want to deny that there is *some* connection between what it is right to do and what it is of value. The virtue of beneficence, for example, is centrally concerned with promoting valuable states of affairs. But for the virtue ethicist, the relation of right action and valuable state of affairs is complicated and depends on which virtues are at stake. This, too, will place a central role in the argument to follow.

4. Moral Risk and the Virtues of Beneficence and Perfectionism

According to many virtue ethicists, an active life of virtue is the greatest of goods. This is first of all to say that one ought always to act in accord with virtue.[9] What exactly this means is determined, in turn, by what particular virtues and goods are relevant to the situation that one faces. If more than one virtue is relevant to the situation, and if their demands can conflict, there will not be an "all things considered" ought in regard to the demands of each particular virtue. What one ought to do, all things considered, will instead have to somehow balance the demands of the different virtues. But even so, we still need to begin by asking what particular goods and virtues are relevant.

In a situation of moral risk, the goods at stake are one's own virtue and the general well-being of others. So what virtues are relevant? Most obviously, one relevant virtue appears to be something like beneficence or sympathy.[10] The virtuous agent considers taking a moral risk not for any trivial reason but because it will (or may) genuinely promote the good of others in significant ways. Since the good of others seems to be what beneficence aims at, it seems that the demands of beneficence are what concern us here. Moreover, the proper (or dominant) mode of moral response for the virtue of beneficence seems to be one of promotion. Beneficence does not aim first and foremost at the agent acting in a beneficent manner. Rather, it aims at actually securing the good for others.

But it is just at this point, I take it, that the problem is thought to arise for the virtue ethicist. The virtue of sympathy or beneficence, it seems, must respond

[8] Cf. Swanton (1995; 2003, esp. chap. 2).

[9] Cf. Swanton (1995; 2003, esp. chap. 2).

[10] Of course, the virtue of beneficence infamously does not appear on Aristotle's list of the virtues. But virtually no one would now deny that it is a virtue; in any event, no contemporary virtue ethicist that I know of does.

to actual values in the world. To act in accord with the virtue of beneficence is to aim to promote genuine goods for the relevant agents in the situation.[11] And if acting virtuously is the greatest of goods, it seems that promoting states of affairs where more of these actions come about is the most beneficent thing one can do. And so if an agent faces a situation in which she can promote the non-moral good of others at the risk of bringing about less virtue, then it seems she should not be allowed to promote the former at the expense of the latter.[12]

But would this advice follow if an active life of virtue is the greatest of goods? It is not obvious that it would. The virtue of beneficence is traditionally understood as the virtue directed at promoting the good of others. It is not a general virtue directed at promoting the amount of goodness in the world. In particular, it is not directed at promoting one's own good. So, even if one's own active life of virtue were more valuable than the non-moral good of others, it would not follow that the virtue of beneficence directed one to promote it in preference to the non-moral good of others.

Rather, it seems that we have here a case of *prima facie* conflict between two different virtues: the virtue of beneficence, which is essentially other-directed, and the virtue, whatever it is, which is concerned with promoting our own moral development. But what virtue is that? It is not clear that the ancient philosophers believed such a virtue existed. When discussing moral development, Aristotle, at least, focuses virtually all his attention on the education of the young. Indeed, he suggests at times that a good upbringing, with the proper habituation, is pivotal if one is to develop into a virtuous adult.[13]

But this Aristotelian emphasis on habituation is not incompatible with the belief that an adult can change or improve her character. What it does imply is that such change will be neither easy nor instantaneous. It will instead require a period of rehabituation.[14] And this seems exactly right. It does seem virtually impossible to become, right this minute, a more caring, courageous, or just person. But over time, we think, people *can* set out on courses of action that would make them more caring, courageous, or just (or more callous, cowardly, or unjust). And insofar as none of us is completely virtuous, people have at

---

[11] I say "promote genuine goods for the relevant agents" very carefully. I say "promote" rather than "maximize," because it is fairly clear that virtue ethics in general ought not to be assumed to be a maximizing theory, not only overall, but even only in regard to the particular virtue of beneficence. And I say "relevant agents," because it is also clear that at least some very influential strains of virtue ethics will want to deny that we need to promote the good of all other agents impartially. Many Aristotelians, for example, will think that we are at least permitted (and perhaps required) to promote the good of those "near and dear" to us in a special way.

[12] As I will argue later, the cases in which beneficence would only promote non-moral goods (and could be known only to promote non-moral goods) are actually quite artificial and *recherché*, given the dependence of *eudaimonia* on non-moral goods. But they are not impossible.

[13] See, e.g., Aristotle, *Nicomachean Ethics*, 1095b3–13, 1103b23–25, and 1152a27–33.

[14] On this point, see Annas (1993, pp. 53–66).

least a *prima facie* obligation to set out on courses of action that would make them more caring, courageous, or just (and not on courses of action that are likely to make them more callous, cowardly, or unjust). The disposition to do this well, virtuously, does seem to be something worth calling a virtue.

Christine Swanton has recently discussed such a virtue; in her words, "there is (paradoxically) a virtue associated with self-improvement" (2004, p. 205). Following Martha Nussbaum, Swanton distinguishes between the thin and thick conception of a virtue. The thin conception simply states the field, or sphere of concern, of the virtue and understands the virtue as being well-disposed toward that field. The thick account of a virtue is an account of what it is to be well-disposed in regard to that particular sphere of concern. "Perfectionism" is the name she gives to the virtue under consideration. On the thin account, perfectionism is simply whatever virtue constitutes being well-disposed toward striving for excellence. This leaves open what the thick account of the virtue will be; rival accounts of "perfectionism" as a virtue will differ on what it takes to be well-disposed toward pursuing excellence.

On one natural reading, an appropriate thick account of perfectionism might seem to tell us simply to imitate the actions of the perfectly virtuous. But as Swanton (2003, 2004) persuasively argues, this account is less attractive than immediate appearances indicate. For someone who is not yet fully virtuous, acting *as* the virtuous person acts, with all the requisite motivations and affective responses, might well be impossible. Given this, simply imitating their actions—doing what they do but not as they would do it—might well result in certain kinds of motivational distortions and bad consequences. Consider first the motivational distortions that may result: drawing on the work of the psychologist Karen Horney, Swanton argues that such a perfectionism can be a vice when it is the mark of what she calls an 'expansionist type.'[15] Such a person identifies himself with his high standards and, on this basis, looks down on others. And such contempt for others, she adds, is really an externalization of self-condemnation on the part of the individual. And one who acts in such a way does not act virtuously, even if his external actions seem to conform to the standards of the perfectly virtuous.

But in addition to these motivational problems, simply imitating the actions of the perfectly virtuous might also lead to very bad results. Swanton notes this but does not give detailed examples of how it is the case. But what I am calling situations of moral risk are perfect examples of this claim, and so provide support for rejecting an account of virtuous perfectionism according to which we ought simply to imitate the actions of the supremely virtuous. Take the case of Nancy again: suppose the virtue of perfectionism were simply the disposition to imitate the actions of the perfectly virtuous. In that case, it

[15] Horney (1970), cited in Swanton (2006), chapter 9.

would seem that Nancy should become a nurse. She already believes that she can do much good by becoming a nurse. Her worry is that she might become less compassionate in doing so. But the perfectly virtuous person would have no such worry: presumably someone perfect in virtue would have the right dispositions to encounter ordinary human suffering without becoming less compassionate. Dealing with such things is, after all, part of what virtue is supposed to equip us to do. So if the virtue of perfectionism told her to simply imitate the actions of the perfectly virtuous, it would seem to imply that she ought to become a nurse and simply hope not to become less compassionate in the process.

But none of this seems right. For one thing, the justification for undertaking the effort appeals to the good one can do. But how much good one can do seems to depend, at least in part, on what kind of person one becomes in the process. If Nancy becomes totally lacking in compassion, she will lose her character; but, just insofar as she does, she might also do less good for her patients. It is true that self-interested or dutiful motives might be adequate for performing some of her duties as a nurse. A self-interested or dutiful nurse might well change bandages with efficiency and give shots on time. But it is far from clear that being a good nurse consists only in performing such actions. Excellent medical care seems also to involve what is referred to (sometimes dismissively) as a "bedside manner." Showing genuine compassion, in other words, is partially constitutive of providing excellent care. To simply ignore this seems to represent a willful blindness, and willful blindness is not virtuous.

So it seems right to say that the virtue of perfectionism "should be seen as a dynamic virtue that is to some extent relative to the imperfections of the agent" (Swanton, 2004, p. 184). What, then, should the virtue of perfectionism tell us about when to take moral risks? In asking this question, I am not yet asking what action would be the right one to do *overall*. The demands of perfectionism and beneficence may conflict, and how to adjudicate that conflict is a question I take up in section 5.[16]

Consider first a case where one is guaranteed that embarking on some course of action will result in one's becoming less compassionate, for example. This is only a case of moral risk in the sense of being a limiting case, since the "risk" in question is a 100% chance of becoming less virtuous. And as such, it seems relatively straightforward what the demands of perfectionism as a virtue would direct one to do: one ought not to embark on the course of action. Even

[16] I am working under the assumption that the demands of the particular virtues can conflict and, in such a situation, we need to determine the course of action that is best overall. I thereby reject the claim—defended by, for example, McDowell (1979)—that any action that is virtuous in some respect is thereby right overall. But I do not think the argument depends in any essential way on my being right about that: if I am not, the argument can be rewritten, substituting "appears to be demanded by beneficence" for "demanded by beneficence," and so forth.

if the immediate action is not, in any other sense, wrong, the virtue of perfectionism is concerned not with the immediate action, but with the relation in which it stands to one's own striving for excellence. If embarking on a course of action is guaranteed to make one worse, then it can hardly be thought to be in accord with the demands of a virtue directed toward making one better.

The interesting cases, of course, are like the one I have focused on here. The agent rightly fears for her virtue, but there is no guarantee that she will become less virtuous. It might seem as if taking such risks, too, would be incompatible with the virtue of perfectionism. How can it be virtuous in some respect to embark on courses of action that we know might make us less virtuous than we already are—or, worse yet, courses of action that might make us vicious?

But on reflection, it cannot be right that perfectionism would never direct us to take such moral risks. It may often be the case that the same course of action might present some risk of making us less virtuous but also some opportunity of making us more virtuous. Someone who decides to become a trauma surgeon might rightfully fear that she would become less compassionate as a result of being exposed to such suffering on a day-to-day basis. But she might also rightfully hope that she would become more compassionate by virtue of her experience. Crucially, this could be true of one and the same person, in one and the same situation. Indeed, for most of us, either one of these outcomes might seem quite possible. And which is more likely to occur might be almost impossible to predict: for one thing, there is simply a limit to the depth of our self-knowledge.

And there is another reason why we should not understand the virtue of perfectionism as telling us never to take moral risks—namely, moral risk is simply pervasive. For persons still imperfect in virtue, no course of action carries absolutely no risk of coarsening their character. Even very admirable actions of self-sacrifice run the risk of puffing us up with pride and moral smugness. Especially admirable acts of self-sacrifice may indeed put us at particular risk of this peculiar moral failing. But it doesn't follow that, in the quest to become better people, we should never perform acts of great self-sacrifice. So the mere risk of a coarsened character cannot be a reason to think an action incompatible with the demands of virtuous perfectionism. If it were, the price would be complete paralysis.

But, of course, not just any moral risk seems justified either. The easiest cases would be those where the risk is real enough, but the possibility of improvement seems negligible or nonexistent. In such situations, it is often the better part of wisdom not to take moral risks. Whether this is so in any case depends not only on the situation, but also on the particular person taking the risk.

So, it seems that the best thing to do is aim for some sort of balance: when one is considering taking a moral risk, one needs to know how likely it is that one will improve, how likely it is that one will deteriorate, and what magnitude

the improvement or deterioration might be. A very large risk seems advisable only if the prospects of improvement are very great. And if the improvement is negligible, then the risk ought not be great.

Of course, none of this gives us an easy answer to the example I sketched earlier. But that is only to be expected. On my account, whether the virtue of perfectionism tells one to take a moral risk depends upon very particular facts about the situation and the person in question. So we simply need more information before we can give a definitive answer. For example: we all know that some of us are less likely to be cowed by social conventions than others. Sometimes this is a particularly moral virtue; we have learned to follow our conscience whatever others may do. Sometimes it is a more basic feature of our personality. We simply don't care very much what others may think of us, if indeed we even notice. Either way, how resilient we are to social pressure is clearly relevant to our choice. Whether medical personnel become hardened to their patients and concerned only about economic profit, for example, may depend partly on whether they work in an environment in which other medical personnel value money more than the good of their patients and, if they do, how strongly they can resist this prevailing norm.

But from a virtue-theoretic perspective, this indeterminacy is in no way worrying. It is part of the canonical understanding of that approach to morality that the "right" or "mean" in action and feeling is relative to the individual, and can only be determined by the person of practical wisdom. No one ought to take inappropriate risks. But what constitutes an inappropriate level of risk for one person may not constitute an inappropriate level of risk for another person. There is simply no answer to the question of whether a particular action would constitute the right "amount" of moral risk in a particular situation, independently of knowing who is performing the action. But likewise, there is simply no answer to the question of whether a course of action is brave rather than foolish or temperate rather than ascetical, independently of knowing who is performing the action. It is only once we know exactly who is proposing to undertake a particular action in a particular circumstance that we can say whether it is virtuous in some respect. And acquiring this kind of knowledge requires a great deal of practical wisdom.[17]

[17] Cf. Aristotle, *EN*, 1106a25–1106b35. This is the point Aristotle is making when he discusses Milo the wrestler and his eating habits. From a strictly numerical perspective, six pounds of food might be a moderate amount. But whether it is the "right" or "mean" amount of food for any particular person depends upon who is eating the food. For most of us, it is probably too much food; for Milo it might be too little. So there is a fact of the matter about how much food one ought to eat, an amount of food that is neither too little nor too much, but that amount is different for different people. The case of virtue is similar.

5. Conflicting Demands: Beneficence and Perfectionism

Thus far, I have argued that there are two virtues whose demands are relevant in a situation of moral risk: beneficence and perfectionism. Beneficence, I suggested, seems mostly to direct one to take such risks.[18] But while perfectionism might direct one to take some moral risks, there are nonetheless some situations in which it seems to direct one to hesitate. Perhaps Nancy is in a situation like this: she is highly unlikely to become anything but less virtuous in becoming a nurse. But even so, she is highly likely to do a lot of good for others. Is she permitted to take the risk? And, if permitted, is she required?

It might seem that this question is easily answered: both beneficence and perfectionism seem directed at promoting various kinds of goods. So, perhaps, if there is an apparent conflict between them, we ought always to choose the demand that will promote the most good. But I have already granted that an active life of virtue is the greatest of goods. So in a case where the demands of beneficence only directly concern the promotion of non-moral goods, why shouldn't the demands of perfectionism always take precedence?

Of course, the suggestion is too simple if understood in one way. That is, it cannot be that just *any* conflict between the demands of two different virtues should be determined by reference to promoting the most overall good. Accepting this line of thought would really be tantamount to rejecting any neo-Aristotelian virtue-ethical approach to ethics out of hand and accepting some kind of consequentialism in its place. In one sense, of course, even neo-Aristotelian versions of virtue ethics can be thought of as teleological theories. Right action is derived from a prior account of either the virtuous person, or the virtues themselves. And both the virtuous person and the virtues themselves are generally defined in terms of some concept of the good: the good life for man, or the particular goods to which the particular virtues respond. So the concept of good does seem, in some sense, prior to the concept of the right. But virtue ethics is not a teleological theory in the sense that Rawls, for example, uses that term; that is, right action is not defined as that which maximizes the good.[19] What is virtuous in any particular case might not maximize the good. To act justly, for example, is not to maximize instances of just behavior. It is to do the just thing, even at the cost of there being fewer just acts overall. So the particular virtues do not necessarily aim at promoting the most good.

Nor does it seem that any *prima facie* conflict between the virtues should be resolved by reference to what would promote the most good. If that were

[18] But it need not always do so: suppose the chance of moral deterioration is real enough, and the good results can only be obtained if one's character is not corrupted. In such a case, beneficence itself might caution one, particularly if one could do significant harm were one's character corrupted.

[19] Rawls (1971, p. 24).

true, then virtually any *prima facie* conflict between justice and beneficence, for example, would have to be resolved in favor of beneficence. For surely such cases of apparent or real conflict arise just when it seems that we would do equal or more good by breaking some constraint on behavior that is generally thought to be a demand of justice. Consider a standard case of this apparent conflict: if we hang an innocent man, we will prevent destructive riots and civil unrest from breaking out. If we do not, such riots will certainly break out and many innocent people will die. The demands of justice are clear: we ought not procure the death of an innocent man. But the demands of beneficence—particularly if we bear some degree of responsibility for the civil safety in virtue of our political or other position—might seem to tell us to go ahead and procure the death of the innocent man. So if we followed a principle which told us to resolve all seeming conflicts between the demands of two different virtues by appeal to which best promoted the most good, it would seem *obvious* that we ought to execute the innocent man. But surely it isn't obvious, which is why this kind of case appears so difficult.[20]

But perhaps it is plausible to think that the demands of perfectionism and beneficence should be resolved by reference to which best promotes the most good. Justice is a special case just insofar as the standard moral response to the demands of justice concerns respecting and not promoting the values in question. But it seems as if both perfectionism and beneficence aim at promoting various goods—beneficence, at the good of others and perfectionism, at one's own excellence. So even if we cannot solve the conflict between the demands of two different virtues *in general* by appeal to the amount of good that would be promoted, perhaps the conflict between two virtues that are concerned primarily with promotion should be adjudicated in such a way.

So, let us suppose, at this point, that we should decide between the apparent demands of these two virtues based on which would best promote the most good. Would it then follow that we should prefer the demands of perfectionism to those of beneficence? It would not. Neither beneficence nor perfectionism aims directly at the greatest good. The greatest good is neither virtue nor non-moral goods, such as health. The greatest good is an active life of virtue. And for an active life of virtue, at least on the Aristotelian view, one needs both virtue and non-moral goods. So it would be wrong to say that the demands of perfectionism ought always take precedence over the demands of beneficence because the former aims at the highest good. Virtue, as a state of character, is not the highest good. It is the active exercise of virtue that is the highest good.

Moreover, the virtue ethicist is not, in any way, committed to the claim that a virtuous individual should view her own virtuous life as, objectively speaking, more valuable than the virtuous life of others. It is simply that I have

[20] Cf. Williams's (1973) famous discussion of Jim and the Indians.

a different relation to my own virtue than I have to others. As I pointed out earlier, the virtue ethicist does admit that there are agent-centered restrictions. But this is not because she thinks that it is objectively worse for her to kill one than it would be for two others to kill two. It is because she bears a different relation to the killing in the first case: she is responsible for whether she kills, and is not responsible for whether others kill. But in this case, we are not talking about an agent-centered restriction. It is part of the definition of a moral risk that the agent is not considering an action that is, itself, morally forbidden (or even problematic). So given this, and given the assumption that we are trying to promote the highest good, there is no reason to give preference to promoting one's own active life of virtue.

For both others and for us, the greatest good is an active life of virtue. But there is an important difference in how we can and should promote that greatest good. In our own case, we can and should promote it by promoting our own virtue. Of course, we need also to secure a decent amount of non-moral goods if we are to develop and exercise our virtue. But once we have secured this decent minimum, our attention rightly turns to developing our own virtue. So, for most of us, the best way to pursue the greatest good *for ourselves* is to exercise the virtue of perfectionism.

But when promoting the greatest good *for others*, the situation is quite different. We should not overestimate our ability to directly "promote" virtue in others. We cannot make people choose rightly, for the sake of the noble. Nor can we make them feel compassion when others are ill, or joy when they succeed. Such attempts at "character education" are notoriously difficult to pull off, and often counterproductive.

This does not mean that we are never justified in trying to promote others' good character. Certainly, parents are often obliged to aim at promoting the virtue of their children. And Karen Stohr has argued persuasively that the virtue ethicist, unlike Kant, should hold that it is sometimes legitimate to aim at promoting the virtue (and not merely the happiness) even of adults who are fully capable of setting their own ends. But she also rightly recognizes that this is very difficult to do well, and needs to be undertaken with great care and humility.[21] So in many, perhaps even most, situations it will not be appropriate to try to directly promote the virtue of others.

By contrast, it is much less problematic to promote the non-moral good of others, by providing for their health, safety, and education, for example. And if Aristotle is right, these non-moral goods are essential to the highest good: we need them both to develop virtue and to exercise it. So if we are aiming to promote the highest good—the active life of virtue—it is perfectly

[21] Stohr (2009).

legitimate to do this by providing essential constituents of that good—namely, non-moral goods.

Given all of this, there do seem to be situations in which the demands of beneficence will take precedence over the demands of perfectionism just insofar as acting in accord with beneficence would better promote the existence of lives in which virtue is actively exercised. Take Nancy's case, for example. If the need for qualified medical personnel is especially dire, then the non-moral goods she will promote by becoming a nurse are very fundamental indeed. If any non-moral goods are a necessary condition of the exercise of virtue, certainly continuing life and health are good candidates. By contrast, the virtue that Nancy will sacrifice by becoming a nurse is less fundamental. Becoming less compassionate is never good. But unless one holds to an extremely strict doctrine of the unity of the virtues, being less than perfectly compassionate will not mean that Nancy ceases to be virtuous in any way. This could be the case in two different ways. First, if one holds to a threshold conception of virtue, Nancy might still be fairly compassionate without being perfectly compassionate.[22] Second, even if Nancy ceases to be compassionate altogether, she might not cease to be just, prudential, and courageous. Given this, Nancy will still be able to live a life in which she actively exercises virtue. She will thus be able to realize the highest good available to human beings, albeit only imperfectly and in part. But if people are dying for lack of medical care, such a possibility will be entirely foreclosed to them. In such a case, I hold that Nancy ought to take the moral risk in question.

6. Two Objections to the Account, with Replies

Objection 1: It seems that the relation of perfectionism to the highest good is conceptually closer than the relation between beneficence and the highest good. Perfectionism aims at developing virtue, which finds its point in the active exercise of virtue. But acquiring non-moral goods need not find its point in the active exercise of virtue. Indeed, in any particular case, fulfilling the demands of beneficence might not contribute at all to promoting the highest good. Because our control over others' virtue is extremely limited, there is no guarantee they will use the non-moral goods we provide in a virtuous, or even neutral, way. Hence, if our aim is to promote the highest good, we ought to fulfill the demands of perfectionism rather than beneficence.

Reply 1: While it may be true that the relation of perfectionism to the highest good is conceptually closer than the relation between beneficence and the highest good, it does not follow that the former is always more likely to

[22] For a "threshold" conception of the possession of virtue, see Russell (2012).

promote the highest good. Conceptual closeness is not the same as empirical likelihood of success. Indirect means of promoting some good can be less, more, or equally effective than direct ones depending upon the circumstances.

All I am arguing is that there are some circumstances in which the demands of beneficence should take precedence over the demands of perfectionism. And surely there are cases in which it is extremely likely that providing for others' material needs will promote the highest good by allowing them to live an active life of virtue. That it does so in an indirect manner is beside the point. Likewise, there are cases in which it is only marginally likely that directly trying to improve our own character will actually improve it and so increase our ability to live an active life of virtue. And in this case, the directness of the intention will do nothing to mitigate its ineffectiveness.

Objection 2: Surely we ought sometimes to relieve the pain of those experiencing great suffering *just because they are experiencing great suffering*. But on my account, it is hard to see how this can be so. Rather, it appears that we will be justified in taking a moral risk for someone experiencing great suffering *only because that would prevent them from living an active life of virtue*. This seems to suggest that we value their life and health only instrumentally, and not for its own sake. From this follows, moreover, that we could never take a moral risk to relieve the suffering of the truly vicious. Preserving the life and health of the vicious person will do nothing to increase the highest good, and a decrease in our own virtue will subtract from it.

Reply 2: This worry, while important, relies largely upon a mistaken and overly moralized view about the relation between non-moral goods and virtue. To say someone is virtuous is, in great part, to say that they respond appropriately to some range of goods that themselves are not characterized morally. The temperate person uses food, drink, and sex appropriately. A large part of being a just person is using material resources appropriately. The courageous person handles fear and confidence appropriately. The relation between non-moral goods and the highest good is therefore not an instrumental one, but a constitutive one. So even if we act to promote the highest good, it does not follow that we do not care about others' life and health for its own sake. Rather, we care both that they are alive and healthy and that they use their life and health well, so that their lives goes well, for their own sakes. Insofar as great suffering undermines the life and health of people, it is therefore not merely instrumentally disvaluable but constitutively so.

However, it does follow from my account that one would not be justified in risking her own virtue to promote the non-moral goods of those whom one knows to be seriously and irredeemably vicious. But it is not clear this is a problem. Practically speaking, this worry could arise only in the rarest of circumstances. Unless one fears that the vast majority of patients whom doctors and nurses treat live seriously and irredeemably vicious lives, for example, it would not give Nancy a reason not to become a nurse. So again, the most

this objection could show is that there are *particular* cases in which it would be wrong to risk one's virtue in order to promote others' non-moral good; it would not show that it would be wrong is *all* such cases. Moreover, consider what a case would have to look like for this objection to arise. Perhaps we can imagine, for example, someone being called upon to provide medical care to an irredeemably vicious and unjust army of oppressors. Supposing that providing the care would pose a risk to her character, it is at least not obvious that she should provide the care. So the account does not have radically counterintuitive results in these sorts of cases.

7. Conclusion

According to virtually all forms of classical virtue theory, an active life of virtue is the greatest of goods. What I have argued in this chapter is that this claim is not only compatible with taking moral risks; in certain circumstances, it may positively require it. But the question of whether and when to take a moral risk arises just insofar as an agent's virtue is not yet perfect and hence fragile to some degree. It is thus one example of our need for what might be called a non-ideal virtue theory. This is an area of inquiry that has received little if any attention from virtue theorists.[23] But in a world where most of us are not yet perfect in virtue, it is well worth pondering.[24]

References

Adams, R. M. 2006. *A Theory of Virtue*. Oxford: Oxford University Press.
Annas, J. 1993. *The Morality of Happiness*. Oxford: Oxford University Press.
Anscombe, G. E. M. 1958. Modern Moral Philosophy. *Philosophy* 33: 1–19.
Aristotle. 1999. *Nicomachean Ethics*. 2nd ed. (trans: Irwin, T.). Indianapolis, IN: Hackett.
Cicero. 2001. *On Moral Ends* (trans: Woolf, R.). Cambridge: Cambridge University Press.
Garcia, J. L. A. 1987. Goods and Evils. *Philosophy and Phenomenological Research* 47: 385–412.

[23] One important exception is Robert Adams (2006), who argues persuasively that virtue is almost always fragmentary and fragile, but is genuine virtue nonetheless. My interest is in the normative implications of this claim.

[24] I presented versions of this chapter at the University of Tennessee, Miami University of Ohio, and a conference sponsored by the Character Project at Wake Forest University. I am grateful to each of these audiences for their feedback. In particular, let me thank Thomas Hurka, Glen Pettigrove, and Nancy Snow. I am also grateful to my colleagues Talbot Brewer and Trenton Merricks for their comments on an early draft, and to Christian Miller for extremely helpful comments on the penultimate draft. This publication was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Garcia, J. L. A. 1990. The Primacy of the Virtuous. *Philosophia* 20: 69–91.
Horney, K. 1970. *Neurosis and Human Growth*. New York: Norton.
Hurka, T. 1998. How Great a Good Is Virtue? *Journal of Philosophy* 95: 181–203.
Hurka, T. 2001. *Virtue, Vice, and Value*. Oxford: Oxford University Press.
McDowell, J. 1979. Virtue and Reason. *The Monist* 62: 331–50.
Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Russell, D. 2012. *Practical Intelligence and the Virtues*. Oxford: Oxford University Press.
Stohr, K. 2009. Minding Others' Business. *Pacific Philosophical Quarterly* 90: 116–39.
Swanton, C. 1995. Profiles of the Virtues. *Pacific Philosophical Quarterly* 76: 47–72.
Swanton, C. 2003. *Virtue Ethics: A Pluralistic Approach*. Oxford: Oxford University Press.
Swanton, C. 2004. Satisficing and Perfectionism in Virtue Ethics. In M. Byron, ed. *Satisficing and Maximizing: Moral Theorists on Practical Reason*, 176–89. Cambridge: Cambridge University Press.
Wallace, J. 2013. *The View from Here*. Oxford: Oxford University Press.
Walzer, M. 1973. Political Action: The Problem of Dirty Hands. *Philosophy and Public Affairs* 2: 160–80.
Williams, B. 1973. A Critique of Utilitarianism. In B. Williams and J.J.C. Smart, eds., *Utilitarianism: For and Against*, 77-150. Cambridge: Cambridge University Press.
Williams, B. 1976. Moral Luck. *Proceedings of the Aristotelian Society*. Supplementary volume L: 115–35.

10 }

Dispositions, Character, and the Value of Acts

Bradford Cokelet

Thanks to the virtue ethics movement, moral philosophers now agree that character and virtue are vital topics. Most would agree, for example, that a complete ethical theory must include accounts of moral psychology, character evaluation, moral education, and moral improvement. But this agreement on the importance of having what we might call a 'character module' in one's theory has not led to any sort of general conversion to virtue ethics, understood as a distinctive approach to ethical theorizing as a whole. On the contrary, most moral philosophers accept that we need a character module, but remain perplexed by virtue ethical proposals to assign character (or character evaluations) a more central or prominent role in our ethical theories.

One flashpoint for the dispute is the *core virtue ethical thesis* that an agent's character can and often does affect the ethical value or deontic status of her acts. If this thesis is false, as a growing number of skeptics contend, then virtue ethical proposals to give character a central or prominent role are effectively hobbled. If character does not affect the value or deontic status of acts, then a theory's character module shouldn't have direct bearing on its action assessment module, and as Crisp (2014) contends, it would be hard to see how virtue ethics could constitute "a third method of ethics"—that is, a satisfying alternative to various consequentialist and deontological views.

In what follows, I defend the core virtue ethical thesis against the counterexample-based attacks marshaled by Hurka (2006), Markovits (2010), and Crisp (2014). In short, I argue that the skeptical attacks on the core virtue ethical thesis hit home if we adopt a reliable trait conception of character, but that, by adopting a personal values conception of character, virtue ethicists can rebuff skepticism about their core thesis.

My discussion proceeds as follows. After some introductory comments (section 1), I discuss the reliable trait and personal values conceptions of character and explain how the personal values conception allows us to better

defend the core virtue ethical thesis (section 2). I then develop and defend the personal values approach by rebuffing several skeptical attacks (section 3) and by sketching accounts of diachronic personal character and aretaic character assessment (sections 4–5).

1. Preliminaries

Before getting to the details, some general remarks about my argument are in order. First, as Hurka (2006) rightly insists, any convincing defense of the core virtue ethical thesis must recognize the distinction between an agent's dispositions and the occurrent motives, intentions, and emotions with which she performs some particular act. To see the import of this distinction consider a simpler case involving an inanimate object: a fragile vase is one that is disposed to break if hit with a rock. We can appeal to the fact that it is so disposed in order to *predict* that it will break before a rock (that someone threw) hits it, and, later, we can appeal to the vase's fragile disposition in order to *explain* why it broke. Analogously, psychological and behavioral dispositions allow us to predict and explain an agent's occurrent mental states and behavior. Consider Ian, who has an anger issue and is disposed to fly off the handle if he thinks his authority is being publically questioned. By appeal to this dispositional fact about his character, we can predict, before Sue questions his decisions at the next office meeting, that he will fly off the handle and explain why he flew off the handle once he does.[1]

With the distinction between dispositions and occurrent mental states and behavior in hand, the skeptics launch their attack on the core virtue ethical thesis. The thesis states that character can and often does affect the ethical value or deontic status of acts; and since an agent's character is determined by her psychological and behavioral dispositions, the thesis is true only if such dispositions can and often do affect the ethical value or deontic status of acts. And the last mentioned claim is the one that the skeptics reject: they grant that an agent's occurrent psychological states and behavior[2] can and often do affect the value of her acts, but they deny that the dispositions that are constitutive of the agent's character can have any analogous effect.

For the sake of argument, I am going to simply grant that an agent's occurrent states do not logically or metaphysically depend on more general

[1] I remain agnostic about questions about whether psychological dispositions are in some sense reducible to categorical properties of the person's brain or body and about whether agents have some form of free will that makes it impossible to predict with certainty what they will do.

[2] Nothing in my discussion hinges on the talk of states or the list of sample states I give. For example, my arguments could be recast in terms of the reasons for which an agent occurrently acts or in terms of occurrent motivating stances instead of states.

dispositional facts about her character. For example, I will grant that a person who is reliably and characteristically disposed to be cold-hearted could on some specific occasion feel and act out of great sympathy or empathy. So my thesis here is that dispositional character, understood as something that outstrips the occurrent motives, intentions, and emotions with which an agent acts in a particular case, can and often does affect the ethical value of that particular act.

Next, I should mention that I will focus on (i) claims about the virtue, worth, and moral goodness or badness of acts and leave aside (ii) claims about the moral rightness and wrongness of acts and about whether they are obligatory, permissible, supererogatory, suberogatory, and so on. So while the core virtue ethical thesis is that an agent's character can and often does affect the ethical value *or* deontic status of her acts, I will only be discussing and defending the more restricted thesis that an agent's character can and often does affect the ethical value of her acts. I focus on this restricted thesis for three reasons. First, there are virtue ethicists who follow Anscombe (1958) in thinking that we should abandon the deontic moral concepts (e.g., *morally right, morally wrong*, and *morally obligatory*), and the viability of any such non-deontic virtue ethical theory depends on the truth of the restricted thesis. Second, the skeptics to whom I am responding all attack the restricted thesis; they argue that dispositional character does not affect the virtue, worth, or moral goodness of acts. Finally, the question of whether dispositional character can affect the deontic status of an agent's act raises all the issues we will discuss in reference to the restricted thesis, but it also raises quite a few more. For example, as mentioned, skeptics about dispositional character affecting the ethical value of acts grant that an agent's *occurrent* motives and emotions can affect the virtue, worth, or moral goodness of her acts. They only question whether dispositional character can do the same. But when we turn to questions about the deontic status of acts, things are different; some skeptics about virtue ethics deny that even occurrent motives, emotions, and intentions can affect the rightness of acts, so defending the claim that character affects the deontic status of acts is simply a more complicated business than defending the claim that character affects the ethical value of acts.[3] For all three reasons, I will focus on the restricted thesis about character affecting the ethical value of acts and leave questions about whether we should abandon the deontic moral concepts or provide a virtue ethical account of them for another time.

With these preliminaries out of the way, I turn to my main argument.

---

[3] For relevant discussion of whether motives can affect the deontic status of acts, see Slote (2003); Sverdlik (2011); and Zyl (2012).

2. Two Conceptions of Dispositional Character

To determine whether dispositional character can affect the value of acts, we have to first settle on a conception of character and say something about which dispositions we have in mind when we claim that an agent's character affects the value of her acts. Generally speaking, I think skeptics about the core virtue ethical thesis have assumed a reliable trait conception of character. On this view, an agent's character is constituted by her dispositions to think, act, feel, and navigate her environment in various virtuous and vicious ways, and her character is evaluated by assessing how reliably she is disposed to act, think, and feel virtuously or viciously in various contexts. For example, proponents of the reliable trait conception would assess how characteristically courageous Dan is by figuring out (a) how much courage Dan actually manifests in his occurrent psychological states and behavior and (b) whether he is disposed to behave courageously, or not, in various situations he could face. Because they are concerned with reliability, they would hold that Dan is more courageous if he disposed to always act courageously on the battlefield than if he is disposed to sometimes exhibit courage and sometimes exhibit cowardice there; and because they are concerned with *global* reliability, they would hold that he is more courageous still if he is disposed to act well in all fearful circumstances, than if he is disposed to act well in the face of justifiable fear on the battlefield but not in his personal life (e.g., when facing fears of intimacy and commitment).

Given a reliable trait conception of character, the core virtue ethical thesis is quite naturally understood as the thesis that variations in the global reliability of an agent's dispositions to embody virtue and vice can and often do affect the value of her acts. So if we take some act an agent performs, the thesis predicts that the act would be better if the agent were more reliably disposed to virtue and that it would be worse if the agent were more reliably disposed to vice. And this is the view that the skeptical arguments attack.

In order to understand the skeptical arguments, it is important to keep in mind that they target the claim that there is some logical, conceptual, or metaphysical connection between variations in dispositional virtue and the ethical value of acts. The skeptics can target that claim while granting that variations in an agent's dispositional reliability to instantiate virtues or vices may be of epistemic importance for act evaluation. For example, consider someone who is reliably disposed to be selfish and heartless. If you are told by a third party that the characteristically selfish and heartless person went out of his way to help a stranger in need, you might find this hard to believe and your knowledge of the person's vicious character might help justify your skepticism. But as the skeptics about the core virtue ethical thesis insist, it may nonetheless be *possible* for the characteristically selfish and heartless person to act in an out-of-character way; and they go on to ask why such out-of-character acts are

less valuable or good than analogous ones that are performed by people with better character. More specifically, the skeptics contend that someone with reliably vicious dispositions could—at least in principle[4]—act with occurrent emotions and intentions of the same type that a virtuous person would act with in a situation of the same type, and they then appeal to examples and claim that such out-of-character (but well motivated) acts are just as virtuous, worthy, and valuable as similar acts performed by people with better dispositional character.

Take Hurka's example of the soldier who dives on a grenade in order to save his comrades and who does so without any ulterior motive, such as a desire for social approbation. Since the soldier knowingly risks his life for a commendable reason, Hurka concludes that his act is courageous. In addition, he claims that its degree of courage and worth (praiseworthiness) is not a function of further facts about how the agent is *disposed* to act, think, or feel in other situations. He writes:

> [I]magine that a military committee is considering whether to give a soldier a medal for bravery. Would they say, 'We know he threw himself on a grenade despite knowing it would cost him his life and in order to save the lives of his comrades. But we cannot give him a medal for bravery because we do not know whether his act issued from a stable disposition or was, on the contrary, out of character'? They would say no such thing, and they would be obnoxious if they did (Hurka 2006, 71–72).

Now, I agree with Hurka that cases like this cast doubt on reliable trait interpretations of the core virtue ethical thesis, but I also think there is an alternative interpretation that is intuitively appealing and that avoids Hurka's attack.

In short, we just need to replace the reliable trait conception of character with a personal values conception. According to a personal values conception, an agent's character is primarily a function of her values and the way those are embodied (or not) in her deliberations, propositional attitudes, feelings, acts, an so on. Let me say a bit more about values, in order to fill out the content of the view.[5] The first thing to mention here is that agents can value a wide variety of objects—for example, relationships, people, and projects—and that they can also value living up to various standards or ideals—for example, aesthetic, ethical, or religious ones. Second, I will assume that valuing something involves

---

[4] Some virtue ethicists fail to clearly distinguish these logical, metaphysical, and epistemic issues, and I believe many arguments for the core virtue ethical thesis are weakened if not undercut by the failure to make the distinction. It may be that there is a logical or metaphysical connection between acting with certain emotions and intentions in a single case and having certain dispositions, but this needs to be established, not assumed. For the sake of argument here, I grant that there is no relevant connection.

[5] My thinking about how to sketch the personal values model is indebted to Scheffler (2011).

integrated cognitive, conative, and deliberative dispositions. For example, if Jane values her friendship with Dan, then she believes it is good or important, is disposed to feel sad if the friendship falls apart, is disposed to avoid threats to the friendship, and is disposed to take such threats into account when deliberating about what to do. Third, agents can embody or live up to their values to a greater or lesser degree. Roughly, I hold that when someone fails to fully live up to or embody her values she values the thing cognitively, but deliberates, feels, acts, and so on, in ways that are out of line with her evaluative beliefs. Fourth, agents usually value many objects, value some more than others, and can harbor inconsistent or conflicting values. Finally, fifth, when we make comparative judgments about how much someone values an object, we are assessing how well she embodies or lives up to her values, not just assessing her evaluative beliefs about the object; and how much someone values something depends on her dispositions, not just her occurrent attitudes and behavior in one instance. For example, how much Jane values her friendship with Dan is underdetermined by her sincerely judging on one occasion that it is more valuable (or important to her) than her career advancement. If she actually makes this judgment but is nonetheless disposed to change her judgment and choose, without compunction or regret, her career over the friendship in conflict cases, then she does not actually value it as much as her isolated judgment suggests.

We would need to do a lot more work to develop and defend a full-blown personal values conception of character, but for my purposes, the very rough sketch of valuing just given will do. What we need to do next is note how the personal values conception can be put to use in interpreting the core virtue ethical thesis and how the personal values interpretation of the thesis differs from the reliable trait interpretation. Now, the core virtue ethical thesis tells us that character can and often does affect the value of acts, so if we adopt the personal values conception, then it will entail that the value of an agent's act can and often does vary with changes in her dispositional values. Put otherwise: for some acts that are actually performed, the thesis predicts that they would be better if the agent had had better dispositional values and that they would have been worse if the agent had had worse dispositional values. The exact content of this prediction depends, of course, on the standard one adopts in order to ground claims about someone having better or worse values, but even without introducing any such standard we can illustrate the view, and its appeal, by returning to Hurka's soldier example and considering a specific way in which the soldier's values might be flawed.

The personal values interpretation of the core virtue ethical thesis implies that the soldier's selfless act will vary in courage and worth when the quality of his dispositional values varies. To pick out some relevant values, I suggest we attend more carefully to the natural assumptions we make when we think his act was courageous. As mentioned before, we are assuming that the soldier jumps on the grenade in order to save his comrades, despite knowing that it

may cost him his life; and I agree that this makes the act courageous. But note that the description given so far, which focuses on the agent's occurrent intentions or motives, does not tell us anything about how much the agent values his life or well-being. It is possible, for example, that the man does not value his life as he should—he may be disposed to think his life is nearly valueless, to sacrifice it or risk it for various minor goods, to not care when others put his life at risk, and so on. And reflection on this case reveals just what the core virtue ethical thesis predicts: the degree of virtue, worth, and value of the soldier's heroic act of jumping on the grenade (to save his comrades) is greater if he aptly values his life than if he does not. Other things being equal, it is more courageous, admirable, and good to knowingly risk something of great personal value than it is to risk something that is, in fact, of great value but that you erroneously take to be a minor good.

We have just seen that the personal values interpretation of the core virtue ethical thesis has plausible implications for Hurka's soldier case, and I will shortly turn to some further skeptical cases. Like Hurka's, the other skeptical cases involve interactions between strangers who are in what we might call "dire straights," and, although I do think the strategy just deployed on the first Hurka case can be applied in many cases of this type, I want to straightforwardly admit that in *some* extra-ordinary cases involving strangers an agent's personal values will have little to no affect on the value of his acts. Admitting this is, however, no embarrassment for the virtue ethicist because it is compatible with the truth of the core virtue ethical thesis; we can admit that dispositional values have no affect on acts in some cases while maintaining that they can and often do affect the value of acts more generally. For example, the virtue ethicists can point out that dispositions affect the value of acts both in *some* extra-ordinary cases, such as Hurka's, and, as I will now argue, in numerous ordinary cases involving people who are substantively related.

To see the pervasive influence of dispositional values on the value of acts in ordinary circumstances, we need only turn to cases in which the value and worth of an agent's act is a function of its interpersonal meaning.[6] The basic idea is that when people are substantively related to us—that is, they are our friends, parents, bosses, and so on, and not just strangers—we care a lot about what they think and feel about us and how much they value us and our relationships with them. By extension, the acts that these people perform have interpersonal meaning to the extent that they embody, communicate, or signify what these people think and feel about us and how much they value us and our relationships.

Take Randy, who shows up at his son's soccer game to support him, and who feels love and pride as he watches his son from the sidelines. Showing up

[6] My thoughts about interpersonal meaning are influenced by Scanlon (2010).

to watch is in this case a good, kind thing to do. Randy's act speaks well of him, and both he and his act are worthy of praise. But, as the personal values view suggests, the ethical value of Randy's act will nonetheless wax and wane with the quality of his dispositional values. For example, we might naturally assume that Randy's act is in character in the sense that his motives, intentions, and emotions in this one case reflect his personal values—we might assume that he values his son and their relationship and that for this reason he is disposed to support and love him in various situations. But, of course, it is also possible that Randy is acting on out-of-character intentions when he comes to his son's game. It is possible that, while he genuinely intends to support his son on this one occasion, he does not value his son and their relationship very much and for this reason he is not reliablely disposed to love and support him. For example, if Randy's friends had invited him to watch the ball game and get drunk at the bar, he would never have showed up at the game, and he would have felt no regret or guilt about not showing up; if asked he would say that "quality time" with kids is overrated and that he would prefer to party with friends, or even to put in some overtime at work. The point is that the value and worth of Randy's showing up (for good reasons) will wax and wane with his overall values; his act of showing up will "mean less" and be less good and praiseworthy, if he is a bad dad than if he is a good one. And this is just what the core virtue ethical thesis predicts on the personal values interpretation.

Before moving on to consider some further skeptical cases, I want to highlight the difference between the personal values interpretation and the reliable trait one. To do so, I want to explain why the moves I have just made in response to Hurka's skeptical case are not available to us if we adopt the reliable trait conception of character instead of the personal values conception. It might seem that proponents of the reliable trait conception could point out that if Hurka's soldier does not value his life as much as he should, then he will have the dispositional vice of imprudence, and that if Randy is a bad dad that is because he has the dispositional vices of disloyalty and heartlessness. Now, in general, I think that there is something right about this form of response. In particular, I agree with the dispositional vice claims just made. But the important point for our purposes is that the response will not save the reliable trait approach. Let me explain.

The reliable trait conception I sketched at the outset emphasizes the importance of being globally virtuous or vicious, and this suggests, for example, that Randy's act would have more value if he exhibited loyalty in a variety of contexts than if he did not; and this is a mistake. The value of his act may well be affected by how loyal and sympathetic he is disposed to be *toward his son*, but that is because his degree of dispositional loyalty and sympathy in this bounded context embodies, communicates, or signifies how much he values his son and their relationship. By the same token, variations in Randy's *global* loyalty or sympathy will not be reliably correlated with variations in

the value of his act, and that is because variation in global loyalty and sympathy may have no substantive connection to how much he values his son and their relationship. So while variations in reliable character may well affect the value of acts, this is true only when (and because) the variations in reliable character reflect variations in the personal values that are relevant in the local context. And this is just what the reliable trait conception of character fails to adequately acknowledge, and what the personal values conception is designed to highlight.

3. Three More Skeptical Cases

I have argued that the personal values version of the core virtue ethical thesis yields intuitive results in ordinary cases and that, unlike the reliable trait version of the thesis, it avoids Hurka's soldier counterexample. In this section, I want to look at three more cases that skeptics take to cast doubt on the core thesis, and explain how to accommodate them. Consider, first, two cases involving strangers in dire need:

> Ronnie and Reggie are vicious gangsters. They are equally ready to kill, torture, and terrorize people, and do so to the same extent. On one occasion, both of them are confronted by someone whose life is in danger and needs help. For some reason (perhaps her perfume reminds him of a past girlfriend), Ronnie feels sorry for this individual, and finds it easy to help her, thereby acting in an entirely uncharacteristic way. Reggie, as usual, does nothing to help (Crisp 2014: 14).

> If a fanatical dog-lover performs a dangerous rescue operation to save a group of strangers, at great personal risk, should we discount the worthiness of his actions because, had his dog required his heroics at the same time, the dog-lover would have abandoned the strangers? That he would have done so may be a sign of excessive concern for the dog, rather than of too little concern for the strangers—after all, the dog-lover was willing to risk his own life to save theirs. And given that the dog was not present to deflect our hero's attention from the reasons he had to perform the rescue, it seems ungenerous to withhold praise for so admirable an act simply because the dog might have been present (Markovits 2010: 210).

As mentioned, I think that defenders of the core virtue ethical thesis should simply grant that in some cases the agent's dispositions have little or no bearing on the value, degree of virtue, or praiseworthiness of her acts, and I will grant, for the sake of argument, that Crisp's case is one of these. Regarding Markovits's case, however, there are two points to make. First, as in Hurka's soldier case, the value, virtue, and praiseworthiness of the dog lover's actions

would be affected by how much he (dispositionally) values his life. Second, although the dog lover's inordinate concern for his pet may not itself affect the value of his heroic act, it would affect the value of his acts in more ordinary cases involving substantive relationships. For example, imagine that the fanatical dog lover gets a call from a friend who says her father just died and that she could use some support from family and friends. The dog lover says, "I will be right over; as luck would have it, my dog Muffy has stopped getting lonely when I go out, so I am happy to come over and stay as long as you want." This sort of response, and follow-up responses clarifying that the dog lover values Muffy's maximal comfort over her friend's basic emotional welfare, are sure to leave the friend unhappy and wishing for better friends. More pertinently, as the core virtue ethical thesis predicts, the fanatic's action of coming over to comfort his friend (for good reasons) is less good and praiseworthy than the similar action of a better friend—one who values her friend's serious need over her dog's trivial comfort and who is therefore disposed to come over in the case in which her dog would be lonely in her absence. Here, as elsewhere, I think we care about the meaning of the actions of our friends, parents, bosses, and so on, and this meaning is often largely a function of people's personal values, which are reflected in their dispositions to think, act, and feel.

Finally, I want to consider another case that Hurka gives us and that is amenable to a response like the one I gave to his soldier case:

> [I]imagine that your companion stops to give $20 to a homeless person, apparently from concern for that person for her own sake. Do you say, 'That was generous of you' or 'That was generous of you on condition that it issued from a stable disposition to act from similar motives in similar circumstances'? Again surely you say the former. Since your judgement is of virtuousness, it turns on your companion's motives rather than on any external features of her act. If you learn that she was acting only to impress you or some bystanders, you may grant that she acted rightly but will withdraw your attribution of generosity. But that attribution concerns only her current motives, apart from any connection to longer-lasting traits (Hurka 2006: 71).

Now, when first reading this case we naturally assume that the companion values the $20 and that in giving the bill, she is therefore knowingly making a generous sacrifice. But just as the soldier might undervalue his life and that would at least diminish the courage of his action, the companion might undervalue the loss of $20 and that would at least diminish the generosity of her action. For example, if the person has been inducted into a strange religious cult, disvalues having money in her wallet, and is therefore disposed to give her money away to anyone who asks, then her act is less generous than it would be if she values what she gives away, and is therefore disposed to be less openhanded with her money. In each case, she may be proximally moved by

awareness of the plight of the homeless person and give the bill solely in order to help him out, but the degree of generosity in her act hinges on how much she values what she gives away.

4. From Personal Values to Personal Character

My argument so far has focused on the claim that an agent's personal values can affect the value of her acts. My assumption has been that if this claim is granted then we can move rather simply to the virtue ethical thesis that dispositional character, understood as something that outstrips the motives, intentions, and emotions with which an agent acts in a particular case, can and often does affect the ethical value of that particular act. But the skeptics might grant my various claims about personal values affecting the value of acts and attack my assumption that these support the virtue ethical thesis; they could grant that personal values affect the value of acts but question the tie between personal values and dispositions. After all, I granted at the outset that an agent's intentions or reasons for acting are not logically or metaphysically tied to her dispositions, so the skeptic could reasonably demand that I justify the assumption that things are different when we turn to personal values. In Hurka's soldier case, for example, the skeptic could admit that the act's value is a function of both the soldier's intentions or reasons for acting *and* how much he values his life, but then suggest that both of these factors are non-dispositional. Why, in other words, couldn't the soldier's out-of-character courageous action involve both out-of-character intentions and out-of-character personal values?[7]

To assess this challenge, it is important to distinguish two different claims one could make about the connection between personal values and dispositions. The first, extreme claim is that our personal values at a time are logically or metaphysically tied to *persistent* or *characteristic* dispositions in a way that makes it impossible to act from out-of-character values on one occasion. For example, one might claim that it is impossible for the soldier to value his life very much right before jumping on the grenade while being characteristically disposed to inordinately disvalue his life. Now, like the skeptic, I don't see why we should accept this strong view. I do think that in ordinary circumstances we may be justified in ruling out this kind of thing as a practical possibility. Moreover, in many cases, evidence of one's persistent or characteristic personal values will be more accessible than evidence of one's values in a single case, so it may make sense to focus on gathering evidence of persistent values when working up an evaluation of an agent or her acts. But since we can't simply leap from those practical and epistemic points to any strong metaphysical

[7] Thanks to Roger Crisp for pressing this objection.

or logical conclusions, I think we should grant to the skeptic that people can sometimes act from temporary, out-of-character personal values.

Luckily, there is a second, more moderate claim about personal values and their connection to dispositions to think, act, and feel. This is the claim that if someone values something at a time, then he is, at that time, disposed to value it in various ways. Take the soldier. If he happens to momentarily value his life a great deal, then, at that moment, he is disposed to protect his life, refuse to sacrifice it for small causes, feel sad at the prospect of his dying, and so forth. If he is not so disposed, then the claim that he values his life a great deal is simply false. In fact, having those dispositions is part of what it is for him to value his life a great deal (at that time).[8]

To see that these more moderate claims about the logical and metaphysical connection between valuing and being disposed are plausible, we can usefully reflect on a second case. Consider again Jane, who claims she values her friendship with Dan more than her career development. My contention is that this claim is true only if she is, at the time she makes the claim, disposed to judge the friendship more important, to choose it over career advancement, to feel bad if she fails to choose the friendship over her career advancement, and so on. So if she claims to value the friendship more but at that very time is disposed to happily sacrifice Dan's happiness to get ahead, to feel glad about the results, and so forth, then her claim is false. Consider, in this connection, how odd it would be for Jane to say, "I value your friendship more than getting ahead; but, if I have to, I will stab you in the back to get ahead and I won't feel bad about it." This assertion flirts with self-contradiction just as my view predicts.

We have just seen that personal values at a time are themselves dispositional and that we therefore may legitimately move from the premise that personal values affect the value of acts to the core virtue ethical thesis; but the distinction between values at a time and values across time suggests a new line of response that the skeptics could press in response to my argument. The skeptics could grant that while *personal values at a time* can and often do affect the value of acts, the same cannot be said of *personal character*, which is presumably a function of how an agent's values, beliefs, intentions, emotions, behavior, and so forth, change over time. In other words, the skeptics could grant that I have established the personal values version of the core ethical thesis but still express skepticism about the personal character version. Moreover, they could point out that the personal character version of the core thesis is the one that most immediately comes to mind when someone claims that character affects the value of acts, and then argue that if virtue ethicists just stick with the personal values version of the core thesis, and disavow the personal

[8] Of course, sometimes something inhibits a disposition from manifesting itself, so the man could value his life and have these dispositions even if they were never manifest in his life.

character version, then it would be misleading for them to claim that their theories assign character an especially central theoretical role.

To fully assess this new skeptical attack and all the issues it raises, we would need to do two things. First, we would need to develop a full account of what personal character is—for example, by asking whether character is fully determined by an agent's dispositional values or whether it also hinges on her success or failure to embody those in her behavior. Second, we would need to develop a typology of the various ways in which personal character can vary in quality and then discuss which types of variations can influence the value of acts. These are not tasks that I can hope to accomplish in the remaining space, of course, but I think I can say enough about the relevant issues to show that the fallback skeptical view—that is, that variations in quality of personal character *never* affect the value of acts—is false.

5. Personal Character and the Strong Virtue Ethical Thesis

As noted at the end of the previous section, personal character is clearly not just determined by what an agent values at a time; but what else it involves is a complicated matter. In general, I think we can say this much: personal character is partially a function of (a) the personal values that an agent is disposed to harbor and manifest at different times and (b) her dispositions to live up to or embody those values or not. To account for cases in which people fail to live up to their values, I also think we can usefully distinguish between how much someone values something and how much he cares about it. To illustrate the distinction, consider the fact that many people care a lot about what their peers think about them and wish they cared less. Many of us know we care about our peers' approval more than we should and, in specific instances, may find ourselves doing bad things because we care about getting peer approval that is valueless, even by our own lights. So as this case illustrates, caring and valuing are distinct and can come apart; we can care a lot about getting something that we do not value at all.

The distinction between caring and valuing will help us develop a plausible account of personal character in part because it will allow us to explain many cases in which agents don't live up to or embody their values. Take Jim, who does not criticize his boss's racist and sexist remarks even though he values standing up in the face of such remarks more than he values being in his boss's good graces. We can explain this failure and Jim's more general lack of integrity by appeal to the fact that he cares more about being in his bosses good graces than he does about standing up to racism and sexism. In addition, the distinction between caring and valuing will also let us distinguish between people who harbor schizophrenic personal values and people who have more or less stable values but who fail to live up to them because of their weak character.

To further illustrate these claims, consider again Jane, who currently values her friendship with Dan more than her career advancement. It is possible that although Jane really does value Dan's friendship more than her career right now, she is also characteristically prone to value things differently at other times or in other contexts. Perhaps when she is at work and around her peers, she values career advancement more than Dan's friendship, but when she is at home, hanging out with Dan, she values his friendship more than climbing the ladder. In my view, there is no more reason to deny that people can have these kinds of schizophrenic personal values than there is reason to deny that people can have characteristically integrated values but, on odd occasions, harbor out-of-character ones. Schizophrenic Jane's values shift from context to context, but she could develop a more stable set of values and still fail, out of weak character, to reliably choose her friendship over her career. For example, even if she values her friendship more than her career advancement both at home and at work, she may be disposed to give in to pressure from her boss and let Dan down as a friend in various ways. She might just care about her boss's approval, or avoiding his angry outbursts, more than she thinks she should.

With these general remarks about the nature of personal character in mind, I want to turn to questions about personal character assessment. Our main goal in this section is to evaluate the claim that changes in the quality of an agent's dispositional personal character—and not just her personal values at a time—can and often do affect the value of acts; and I think personal character can indeed have this affect because it can affect the symbolic value or personal meaning of an act.

Here is an example: the ethical value of Randy's coming to the ball game to cheer his son on is in part a function of how much he values his relationship with his son and the opportunity to support him. Presumably, Randy's son values his showing up and cheering for him in good part because of the symbolic value of those actions—the actions are valuable because they embody and signify something about how much Randy characteristically values his son and his relationship with him. So if Randy does *not* characteristically value his relationship with his son very much, and just temporarily values it a lot on one occasion, his action of showing up will carry less symbolic value than if he shows up and does characteristically value his son. If Randy for some odd reason temporarily and uncharacteristically values his son and their relationship when he shows up, his son might say, "Sure, right now he is all enthusiastic about supporting me and cheering me on and he is waxing poetic about how important family is; but just wait and see; when tomorrow rolls around things will be back to usual." We can naturally imagine him saying this to explain why he is only mildly pleased to see his Dad yelling in support from the sidelines and why his Dad's actions don't mean as much to him as they would if they were indicative of his deep values. By the same token, he would be more

pleased if his Dad's action *was* symbolic of his characteristic values, and that is presumably because the act would then have more meaning or symbolic value.

If I am right, this example illustrates the general truth that personal character, understood as something that is a function of one's dispositions to value at various times, can and often does affect the value of acts, just as personal values at a time do. So the fallback skeptical line of response to my earlier argument about the significance of personal values at a time fails. As I have said, a full discussion of the issues in this area would require a more developed account of personal character, but I will have to leave that for another occasion. What I *will* do in the remaining pages is explain why I don't think that the argument that I have just made about the affect of personal character on the value of acts is available to philosophers who adopt the reliable trait conception of character.

To begin, it is worth noting that we may value a friend's disposition to value us (or our friendship) independently of whether she has a more general and reliable trait to value friends (or friendship). By extension, we might value a friend's act of visiting us in the hospital because it symbolizes or expresses her loyalty to us, without thinking that the act symbolizes or expresses a general, reliable trait of being loyal. These points suggest that the personal values interpretation of personal character fits with the foregoing argument in a way that the reliable trait conception does not, but they also draw attention to the fact that our aretaic (roughly virtue and vice) judgments of people and their acts are often tied to bounded, local contexts that are roughly defined by their connections to some roles, projects, ideals, or relationships. A judgment about our friend's loyalty for example might focus on the personal values that she harbors and embodies in the context of our friendship, not in friendships or life in general. Put otherwise, aretaic judgments often rest on assessments of the personal character someone embodies in a *local* context instead of on her more *global* traits. This point is important to note because my examples in this chapter suggest that it is primarily locally bounded aretaic judgments, grounded in facts about an agent's locally relevant personal values, that can justify the claim that facts about aretaic character affect the value of acts.

To further illustrate the way in which we regularly make locally bounded aretaic judgments and the fact that they are grounded in locally relevant personal values, imagine that you are thinking of hiring Jane and the letter of reference form asks me to describe her character. In this context, you are asking me to describe the facts about her behavior, emotions, traits, and so forth, that are indicative of the personal values that you can expect her to harbor and act on (or fail to act on) while at work. On reflection I might describe her as loyal, compassionate, and easygoing. If, by way of contrast, you are thinking of *dating* Jane and ask me to describe her character, I should describe a very different set of values and dispositions that are relevant in this different context. In assessing her as a prospective partner, I might say she is not very loyal or easygoing. Of

course, the work-related judgments about her being loyal and easygoing might still hold true and might be worth communicating if you wanted to enter into a long-term relationship with Jane, but the main point is that her interactions with you as a partner would be liable to be affected by her degree of relationship loyalty in a way that they would not be affected by her on the job loyalty.[9]

If what I have just suggested is correct, then we can agree that an agent's degree of virtue or vice can and often does affect the value of her acts, without accepting the reliable trait interpretation on which degrees of virtue and vice should be understood in terms of the reliability of the relevant traits. If that reliable trait interpretation is accepted, we have to say that the value of Jane's loyally defending Dan's name in the face of malicious gossip should vary depending on how reliable her disposition to be loyal is. And as we have seen, that is just the kind of claim that the skeptics rightly challenge; the value of Jane's sticking up for Dan does not depend, for example, on facts about whether she would betray her country in wartime, or not. But once we adopt the personal values approach to thinking about character and notice that aretaic character assessments are often local, we have a more reasonable way to think about how aretaic character can bear on the value of acts. The basic idea is that we use virtue and vice terms to describe context-relative personal character and that can indeed affect the value of acts. If Jane is characteristically a loyal friend to Dan, then her act of sticking up for Dan in the face of gossip may have more symbolic value or meaning than it would otherwise have; but facts about how loyal she is in other contexts (e.g., wartime) will not have similar significance. In sum, aretaic dispositional facts may in some cases track the value of acts, but that is only because they reflect some relevant personal values, and not because they reflect generally reliable aretaic traits.

Before concluding, let me mention one other case involving symbolic value or meaning that is a bit harder to assess. I think the Randy and Jane cases show that facts about personal character can affect the meaning, and hence value, of acts, but these are cases involving personal relationships and it is no surprise that we care about the symbolic value of the ways that our friends, family members, partners, and so forth, interact with us. I want to now consider a case in which someone's more general ethical character might be thought to affect the value of her interactions with strangers.

Imagine that some stranger, who does not owe you anything, helps you out of a jam, perhaps by paying off some debt you have. You are initially grateful and think the act was very good, but then find out that your benefactor is an unrepentant racist who regularly helps out strangers in need, but only after checking to see whether they are of African descent. Now I can imagine some people thinking

[9] Biographers may aspire to give complete descriptions of personal character, but this just highlights the fact that our ordinary assessments are most often local, not global. If MacIntyre (1999) is right, there might be something bad about our tendency to make local but not global character assessments. I don't think this is the case, but cannot pursue the issue here.

that the help you got in this case was in some way symbolically tainted because of the racist personal character of your benefactor. It seems too strong to say that by accepting the help, you made yourself complicit in his racist agenda or that you became party to his plans, but it is entirely conceivable that one would dislike being associated with the racist by being his beneficiary, and for that reason conclude that it was a bad thing to have gotten his help. This implies, however, that you think that the act of his giving you the money is in some way worse because of the way it symbolically connects you to the racist's immoral character.

I end with the case involving the racist benefactor for several reasons. First, I think it illustrates how unclear the concepts of symbolic value and meaning are. The fact that they are unclear does not mean that they are bankrupt, but I want to admit that to fully rebuff the skeptical attack, a virtue ethicist who adopts the personal values approach needs to clarify the relevant concepts. Second, the racist benefactor case suggests that facts about an agent's ethical character can affect the value of her acts even if these are facts about how they would treat people who are not affected by the actions being judged. The ethical value of your interacting with me can depend on how your personal character disposes you to treat other people in some very different context. Finally, third, this case suggests that although meaning most obviously affects the value of acts in cases involving substantive relationships, it can also affect the value of acts between strangers.

6. Conclusion

I have argued that if virtue ethicists adopt a personal values conception of character instead of a reliable trait conception, then they can plausibly defend the core virtue ethical thesis. In fact, they can rightly maintain that personal character, (local) aretaic character, and personal values at a time can, and often do, affect the value of acts. Of course even if this thesis is true, virtue ethicists will need to do more to explain what they mean when they say that they give dispositional character a central or especially prominent role in their ethical theories and to defend the claim that their approach constitutes a genuine third method of ethics. But I hope that in this chapter I have at least shown that these tasks can be sensibly pursued and that wholesale skepticism about the core virtue ethical thesis is unwarranted.[10]

[10] This paper was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. I benefited from helpful feedback from the audience at the final Character Project conference, and also want to thank Roger Crisp, Tom Hurka, Rachana Kamtekar, Michael Slote, Ben Yelle, Dan Corrigan, Anita Kelly, and, especially, Christian Miller for very helpful comments or discussion.

References

Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33, no. 124: 1–19.

Crisp, R. Forthcoming. "A Third Method of Ethics?" *Philosophy and Phenomenological Research*; currently published on-line as an Early View article. DOI: 10.1111/j.1933-1592.2012.00607.x

Hurka, T. 2006. "Virtuous Act, Virtuous Dispositions." *Analysis* 66: 69–76.

MacIntyre, A. (1999). "Social structures and their threats to moral agency." *Philosophy* 74 (3): 311–29.

Markovits, J. 2010. "Acting for the Right Reasons." *Philosophical Review* 119: 201–42.

Scanlon, T. M. 2010. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap Press.

Scheffler, S. 2011. "Valuing." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by Wallace, Kumar, & Freeman, 23–42. New York: Oxford University Press.

Slote, M. 2003. *Morals from Motives*. New York: Oxford University Press.

Sverdlik, S. 2011. *Motive and Rightness*. New York: Oxford University Press.

Zyl, L. van. 2012. "Review of Steven Sverdlik, *Motive and Rightness*." *Ethics* 122, no. 3: 627–32.

11 }

Exemplarism and Admiration

Linda Zagzebski

1. Introduction

There are at least two distinct ways in which we respond to good lives and other things we call good: we *admire* them or we *desire* them. We admire compassion and bravery; we desire health and friendship. We desire financial resources, but we admire the way some people use those resources. Some lives are admirable, but not desirable. For instance, we admire a life of martyrdom or self-sacrifice, but other things equal, do not desire such a life. Similarly, most of us desire a life that has all the goods of human well-being—health, love and friendship, characteristic human pleasures, financial security, political freedom, and interesting work—but even if such a life is free from vice, we typically would not admire it as much as the more virtuous life that lacks some of the goods of well-being.

I believe that neither admiring nor desiring can be reduced to a more basic attitude toward what we call good. I have been working on a moral theory based on the emotion of admiration. It is a map of the moral domain that derives the good in the sense of the desirable from the good in the sense of the admirable. It also derives the deontic concepts of right act, wrong act, and duty from the admirable. In particular, the theory is a map of the moral domain in which the foundation is exemplars of moral goodness to which we refer directly through the emotion of admiration. I call the theory *exemplarism*.

In constructing a theory in which the admirable is more basic than the desirable, I am reversing the priority Aristotle gives in the *Nicomachean Ethics (NE)*. There he defines the good as the desirable—what everyone desires for its own sake (*NE* 1094a18–20), and he identifies that with *eudaimonia*, a life of well-being or flourishing. Aristotle defines virtues, or admirable traits, as constituents of a life of *eudaimonia* (e.g., Ackrill 1980), or on another interpretation, as means to such a life (e.g., Kraut 1989). Either way, the good in the sense of the desirable is basic, and the good in the sense of the admirable

is conceptually derivative. In contrast, I propose that the good in the sense of the admirable is basic, and the good in the sense of the desirable is conceptually derivative. In the next few sections I will address the nature of admiration, the kinds of admiration, and some psychological studies of admiration, and will then describe a way that admiration can be made foundational in a moral theory. In the last section, I will briefly review the advantages of this approach to designing a moral theory.

2. What Is Admiration?

Admiration is an emotion, and so I will begin with a summary of my view of emotion in general. First, I assume that an emotion has an intentional object. Having an intentional object is roughly what distinguishes an emotion from a sensation or a mood. We fear something, hope for something, pity someone, love someone, feel indignation at some state of affairs, feel sympathy with someone's plight, feel angry at someone, and so on. When I reflect, I can reasonably ask whether my emotion is appropriate for its object. What I fear might or might not be fearful. What I pity might or might not be pitiable. The object of my anger might or might not deserve my anger. For some emotions, we think that the issue of fit is relative to the subject. For instance, if I feel annoyed at someone, it might be fitting for me to feel annoyed even though others do not feel annoyed in the same circumstances, and annoyance would not be fitting for them. Even so, I might later judge that I should not have been annoyed, and that suggests that there is an issue of fit even when what counts as fitting varies from person to person. So in general, emotions can fit or not fit their objects, and we think that an emotion ought to fit its object. I do not assume that this feature applies to all emotions, but I will assume that it applies to admiration. Admiration has an object, and it may or may not fit its object. What we admire may or may not be admirable.

Second, an emotion has an affective component. It has a characteristic "feel." It feels different to admire someone than to love her or to fear her or to be angry at her. Although there can be subconscious emotions, I think that we identify emotions in part by their characteristic feeling when they are conscious. I will argue that there are at least two different kinds of admiration, and it is likely that they do not feel exactly the same way, but I assume that there is a general similarity in the feeling aspect of admiration in its various instances.

Third, I assume that because an emotion has an affective aspect, it is potentially motivating. It can move us to respond by expressing the emotion overtly, or by overt action. Admiration, like other emotions, tends to move us, at least when the object of admiration is a person. In particular, I think that it tends to move us to emulate the admired person in the respect in which the person is admired.

I also believe that in a state of admiration, the object of admiration appears admirable, and I think this is a general feature of emotions. In a state of fear the object appears fearsome; in a state of pity the object appears pitiable. I am not suggesting that an emotion includes a judgment or belief. Something can appear fearsome to the agent when she does not judge that it is fearsome. In fact, she may judge that it is not fearsome. But if something appears fearsome to her when it is not, there is a misstep of some kind—a lack of fit between her faculties and her environment. The faculty or disposition through which something appears to her in her emotional state is misrepresenting the object. Similarly, if someone appears admirable when she is not, there is a misalignment between the world and the emotion disposition operating in that situation, whether or not the emotion is accompanied by the judgment that she is admirable. This feature of appearance in emotion is one explanation for the fact that an emotion may or may not fit its object.

Admiration, then, can be generally described as a state consisting of a characteristic feeling of admiring something or someone who appears admirable. Admiration need not include the judgment that the object of our admiration is admirable, but if we trust our emotion, we will make that judgment. We tend not to make the judgment if we are skeptical of our emotion state, or if we wish to withhold judgment until we have been able to reflect. The fact that we think we can critique our emotion is another reason to think that there is an issue of whether admiration in a given instance fits its object.

My account of emotion includes one other feature that I will mention briefly, but it is not necessary to accept this feature for the purposes of exemplarism. I think that the feeling component of an emotion cannot be detached from the appearance. The feeling of pity would not be *that* feeling were it not directed at something that appears pitiable. The feeling of fear would not be that feeling were it not directed at something that appears fearsome. The feeling of admiration would not be that feeling were it not directed at something that appears admirable. The appearance of admirability in the sense I mean therefore has the admiring feeling attached to it. The object cannot appear admirable in this sense without the admiring feeling directed toward it, and the admiring feeling cannot be directed toward something without the appearance of admirability. I have argued elsewhere that if I am right in my proposal that the cognitive and feeling aspects of emotion cannot be pulled apart, that solves some puzzles about the way we treat moral judgments as both cognitive and motivating.[1] However, it is not necessary to accept this point about emotion for the purposes of exemplarism.

---

[1] In Zagzebski (2003) and (2004), I call concepts that we apply to the objects of emotion—pitiable, fearsome, contemptible, rude, etc.—"thick affective concepts." I think these concepts are not purely descriptive, but are affectively laden. They are not understandable apart from the disposition to have feelings of the relevant kind. I argue in Zagzebski (2003) that these concepts are used in what I call ground-level moral judgments. The "thinning" of these concepts explains why we find moral weakness both puzzling and common.

In other places I have said that the object of admiration can be described most generally as the "imitably attractive" (Zagzebski, 2006, 2010, 2012). A disadvantage of this way of describing the emotion of admiration is that it builds the typical response to the emotion—imitation—into the description of the emotion itself. I do not want to say that the urge to imitate or emulate is a necessary component of the feeling of admiration, but calling attention to the reactive response reminds us of what it feels like to have the emotion to which we typically respond in that way. Saying that in a state of admiration the object appears imitably attractive highlights two features of the feeling of admiration that I want to emphasize: (1) the object appears attractive, not repulsive or evaluatively neutral; (2) the way in which the object is attractive typically gives rise to the urge to imitate or emulate the object if emulation is possible. I think we can also say that in standard cases, admiration is a positive emotion in another sense. It is not only directed at what appears good, but it feels good to have it, although admiration can become a bad feeling, as when it leads to resentment or envy. However, I will not discuss resentment and envy in this chapter.[2]

3. Kinds of Admiration

We may get a clearer idea of what admiration is like by looking at similarities and differences among the different kinds of persons who can elicit our admiration. If I were to make a brief list of persons I admire, it would include the following:

- Marie Sklodowska Curie, the Polish physicist and chemist, who won Nobel Prizes in both physics and chemistry for her pioneering work on radioactivity.
- The Trappist monks of Tibhurine, who continued their ministry to the people in their region amid danger in war-torn Algeria, and were subsequently kidnapped and murdered by rebels.[3]
- St. Catherine of Siena, a fourteenth-century mystic and forceful participant in the political affairs of the time, noted for effectively standing up to popes, most famously when she convinced Pope Gregory XI to return to Rome from Avignon.
- Leopold Socha, a Polish sewer inspector and former criminal, who protected Jews hiding in the sewers of Lvov for fourteen months, first for money, then gradually out of compassion and at great personal risk.[4]

---

[2] I discuss the distortions of admiration in envy and *ressentiment* in "*Exemplarist Virtue Theory*," chapter 2, in progress.

[3] See Kiger, 2003; and Xavier Beauvois's film, *Of Gods and Men*, 2010.

[4] See Chiger, 2008; Marshall, 1990; and Agnieszka Holland's film, *In Darkness*, 2011.

– Arthur Miles, the protagonist of C. P. Snow's novel, *The Search* (1958) Miles is an ambitious scientist doing groundbreaking research in crystallography. At the point of making a major discovery, he finds counter-evidence that he is at first tempted to destroy, but then accepts it and reports that the hypothesis that would have made him famous is false.[5]
– Lorenzo Ghiberti, creator of the breathtaking Baptistry doors in Florence.
– Suzie Valadez, who has devoted her life to ministering to the poorest people of Juarez following a religious awakening that made her want to imitate Christ as perfectly as possible.[6]
– Brian Shaw, said to be the strongest man in the world.[7]
– Confucius, perhaps the first sage in recorded history.
– A woman I know who is impeccably groomed and keeps her home always ready for company, while caring for her husband with Alzheimer's.
– Any man who can sing the last verse of "Walk the Line".

In spite of the wide diversity of the individuals on this list, I think they have something in common that elicits my admiration. Each has a distinctively human power in a high degree of excellence—intellectual or artistic genius, moral leadership, stalwart courage, compassion, open-mindedness and intellectual honesty, the virtues of a hostess and loving spouse, physical strength, the ability to sing as low as Johnny Cash. Obviously, some of these qualities are more important than others. Some of them do not have much to do with what makes a person the person that she is, whereas others are much deeper, integrated into the whole of the person's life, such as helping the poor is for Suzie Valadez. In contrast, the ability to sing low or to lift heavy weights is far from exhausting the qualities of a person, even a person who focuses most of his attention on developing that ability.

My hypothesis is that the main division among the kinds of persons we admire is the division between those who are admirable for a natural excellence and those who are admirable for an acquired excellence. We admire both because both are ways in which humans are excellent, but the emotion directed toward natural talents feels different from the emotion directed toward acquired excellences. It seems to me that admiring genius feels different from admiring generosity or kindness or courage. There is also a difference in the way the two kinds of excellences can be emulated. I can imitate a person with an acquired excellence, but I cannot imitate a natural talent. Nonetheless, people who excel in intellectual or artistic excellence or physical strength improve it

[5] Discussed by Jason Baehr in Baehr 2010, p. 142.
[6] See Colby and Damon, 1992, chapter3.
[7] See Bilger, 2012.

with practice, and it often takes perseverance and courage to develop the talent. Brian Shaw became stronger by lifting weights and other exercises. Lorenzo Ghiberti succeeded in creating panels for the Baptistry doors through perseverance as well as talent. Both Shaw and Ghiberti are emulable because the level of excellence they achieved is partly due to effort and acquired traits that can be imitated. I am a person on the lower end of the scale of physical strength, but even I can lift weights. I do not have Ghiberti's artistic genius, but I can imitate the perseverance he needed for success. However, I doubt that there is anything to imitate in singing low. Either you can do it or you cannot. So my (testable) hypothesis is that we do not try to imitate a purely natural excellence or the natural part of an excellence because it cannot be imitated, but the acquired part can be, and there is a difference in the way it feels to admire natural excellence and the way it feels to admire acquired excellence. The object of admiration for a natural excellence is not imitably attractive.

Another way to look at the difference between admiration for natural talent and admiration for acquired excellence is to consider the opposite emotions. If we feel admiration for an acquired excellence like kindness, we feel contempt for an acquired defect like meanness. Meanness is not simply the absence of kindness. It is an acquired trait that is opposed to kindness. But whereas we feel admiration for a natural talent, we do not feel contempt for a state that is opposed to a natural talent. For one thing, there is no such thing as an anti-talent, a quality opposed to natural talent the way meanness is opposed to kindness. There is, of course, such a thing as the lack of talent, but I do not think people usually feel contempt for it. For the most part, I doubt that we have any emotion at all toward a person who lacks a particular talent. If someone is truly lacking in all talent, we might feel pity, but certainly not contempt. I think, then, that admiration for natural talent and admiration for acquired excellences are two different kinds of admiration. The latter kind of admiration has contempt as its contrary. That is the kind of admiration I am interested in for exemplarism.

Although I think we admire natural intellectual talent in a different way than moral virtues, I think we admire intellectual virtues in the same way we admire the moral virtues. The intellectual honesty and selfless open-mindedness of C. P. Snow's character, Arthur Miles, is admirable in the same way honesty is admirable, and it is admirable in the same way self-sacrificing virtue of any kind is admirable. I think also that closed-mindedness and intellectual dishonesty are contemptible. If so, the main division in kinds of admirability is not the division between intellectual and moral traits. It is the division between the natural and the acquired.

In this section I have proposed that admiration is directed at human excellence of all kinds, but the primary division in the range of the admirable is that between natural and acquired excellences. The two kinds of admiration feel different to the person experiencing them, they have different

contrary emotions, and they differ also in the way in which we respond to them. We can imitate acquired excellence, but not natural talent. These hypotheses can all be empirically tested. I am not suggesting that this division in the kinds of admiration and the admirable is something we know a priori. If the empirical results do not support my contentions, the way exemplarism is applied will need to be modified. But if the contentions are correct, we can focus our attention in constructing the theory on admiration for acquired excellences.

Another hypothesis I proposed in this section is that intellectual virtues are among the acquired excellences, that they are admirable in the sense that is opposed to the contemptible, and that they belong to the class of traits relevant to exemplarist moral theory. A sage like Confucius is someone whose admirability blends the moral and the intellectual, and I think it is morally relevant that Confucius has intellectual virtues. But readers who believe an ethical theory should be restricted to the domain of traits and acts that are good in a distinctively moral way can easily adapt the points I will make about admiration and exemplarism in this chapter to their point of view.

4. Haidt on the Psychology of Admiration

I know of very little work on admiration by psychologists, but an exception is a series of articles authored or coauthored by Jonathan Haidt.[8] Haidt calls admiration an other-praising emotion, arising from witnessing the non-moral excellence of others. Haidt invents the term "elevation" for the other-praising emotion arising from witnessing the moral excellence of others that does not benefit the self.[9] Haidt says he got his idea for the features of elevation from a letter by Thomas Jefferson to a friend on the benefits of reading great works of fiction. Jefferson writes:

> When any . . . act of charity or of gratitude, for instance, is presented either to our sights or to our imagination, we are deeply impressed with its beauty and feel a strong desire in ourselves of doing charitable and grateful acts also. On the contrary when we see or read of any atrocious deed, we are disgusted with its deformity and conceive an abhorrence of vice. Now every emotion of this kind is an exercise of our virtuous

[8] Haidt, 2003; Algoe and Haidt, 2009; Vianello and Haidt, 2010.

[9] Haidt says that the emotion directed toward the moral excellence of another that does benefit the self is gratitude, and he places gratitude in the category of other-praising emotions. I will not discuss gratitude here, but it seems to me it is quite different from emotions focused on praise for another. Jews rescued during WWII felt the other-praising emotion Haidt calls elevation even though they were the beneficiaries of the rescuers' acts. In addition, they felt gratitude, but the gratitude was directed at the personal benefit, not the fact that the benefit was produced by an admirable act.

> dispositions; and dispositions of the mind, like limbs of the body, acquire strength by exercise.

Jefferson goes on to say that fiction generates physical feelings as much as those caused by observing real persons:

> [I ask whether] the fidelity of Nelson, the generosity of Blandford in Marmontel do not dilate [the reader's] breast, and elevate his sentiments as much as any similar incident which real history can furnish? Does he not in fact feel himself a better man while reading them, and privately covenant to copy the fair example?[10]

Haidt uses Jefferson as a basis for his proposal that "elevation" is the name of the emotion we direct toward moral exemplars. That emotion has features that he has confirmed in subsequent small research projects, including interviews in Japan and India. The main components he identifies are the following:

a. It is elicited by acts of charity, gratitude, fidelity, generosity, or any other strong display of virtue.
b. It leads to distinctive physical feelings, including the feeling of dilation or opening in the chest, combined with the feeling that one has been uplifted or 'elevated.'
c. It gives rise to a specific motivation or action tendency: emulation, or the desire to perform the same kind of acts oneself.

Haidt invents the term "elevation" for the emotion he describes, although I think "admiration" is a perfectly good term for the emotion directed toward either morally or non-morally exemplary persons. Algoe and Haidt's studies indicate that both the emotion they call "elevation," directed at exemplars of moral excellence, and the emotion they call "admiration," directed at exemplars of non-moral excellence, such as skill and talent, motivate the subject to emulate the person at whom the emotion is directed in some way, with the difference that admiration for natural talent energizes people to work harder to succeed at their own goals, whereas elevation leads them to emulate the moral goals of the other.[11] I think, then, that in spite of our differences in terminology, Haidt's research shows us some interesting features of admiration in my sense, and confirms my hypothesis on the connection between admiration and emulation.

In addition to terminology, there are some differences between Haidt's view and mine that I want to note. Haidt's studies on elevation focus on the moral excellence of others that does not benefit the self. His examples include

---

[10] Jefferson ([1771] 1975, p. 350), quoted in Haidt 2003. A portion of the passage is also quoted in Algoe and Haidt, 2009.

[11] Algoe and Haidt, 2009, pp.123–24. See especially table 5 and the discussion of study 3.

charity, generosity, fidelity, and gratitude, or any other strong display of virtue. But I can feel elevated by someone's virtue even if I am a beneficiary, as when someone's generosity includes generosity to me. Jews rescued from the Nazis felt elevation for the courage of the rescuers even though they were the beneficiaries of the rescuers' courageous acts (mentioned in note 10). Furthermore, there are intellectual virtues that have the features Haidt identifies in elevation. For instance, I think we feel as elevated by the selfless open-mindedness of Arthur Miles in the C. P. Snow novel I mentioned earlier as by displays of charity, fidelity, and generosity. In addition, there are intellectual virtues that are just like moral virtues except that they apply in the intellectual domain. There is no reason to think that the intellectual humility, the intellectual courage, and the intellectual perseverance of another are less elevating than humility, courage, and perseverance in the "moral" domain. These intellectual virtues benefit others as well as the self, and they motivate us to emulate the person who displays these virtues. However, I do not offer these comments on Haidt's and Algoe's work as significant objections since I assume that their characterization of admiration and elevation is intentionally vague, and even if it is meant to apply to a narrower range of admirable traits and acts for the purpose of the studies they conducted, further research can show whether their conclusions apply to the broader sense of exemplarity I am using here.

Algoe and Haidt do make one claim that I think is importantly different from the view of admiration I am proposing. They suggest that the emotion that is the opposite of elevation is disgust, the feeling of moving down on what Haidt calls the purity scale, and which is elicited by witnessing vicious acts (Algoe and Haidt, 2009, p. 106; the same features are identified in Haidt 2003). This idea seems to be suggested by the quotation from Jefferson that Haidt uses, where Jefferson says that "when we see or read of any atrocious deed, we are disgusted with its deformity and conceive an abhorrence of vice." Haidt thus subsumes the virtue-vice scale to the purity scale, with disgust at the lower end and elevation at the upper end. It seems to me, however, that the purity scale does not coincide with the virtue-vice scale, and I have already said that I think that the emotion opposed to admiration (or elevation in Haidt's terminology) is contempt, not disgust. A person who is brutal or unjust or stingy is contemptible. A person who has sex with a dead chicken (one of Haidt's well-known examples) or eats feces is disgusting. I would not deny that many acts are both contemptible and disgusting, but it seems to me that contempt is the emotion directed toward moral defect.

Let me summarize what I think we can say about admiration so far:

1. Admiration is the emotion of feeling a distinctive way toward a person seen as admirable. The emotion may or may not fit its object. That is, the object may or may not be admirable. But other things

equal, if we trust our emotion of admiration, we will judge that what we admire is admirable.

2. Admiration for acquired human excellences differs from admiration for natural talents. Contempt is the emotion opposed to admiration in the first category, but not the second. The kind of admiration we want for exemplarism is admiration for acquired excellences.
3. The acquired excellences relevant to exemplarism include intellectual virtues.
4. Admiration for acquired traits is often associated with physical feelings, particularly feelings of being uplifted or elevated by witnessing or reading about something admirable (supported by the Haidt studies).
5. The admired person is imitably attractive. That is, admiration for acquired excellences gives rise to the motive to emulate the admired person in the way in which she is admired (also supported by the Haidt studies).

5. Trusting Admiration

Admiration clearly can go wrong, and we know that it cannot always be trusted. I have recently discussed the issue of trusting our emotions at some length, and have argued that basic trust in our emotion dispositions as well as our epistemic dispositions is both rational and inescapable.[12] In both cases we ultimately have no non-circular way to tell that the outputs of these faculties fit their objects. But I argue that it is rational to trust them when they survive conscientious self-reflection, by which I mean reflection on our total set of beliefs, emotions, and other psychic states when we are using them the best we can to make them fit their objects. For instance, we are epistemically conscientious when we do our best to make our beliefs true. We are emotionally conscientious when we do our best to make our emotions fit their objects—pitying the pitful, fearing the fearful, admiring the admirable, etc. Radical skepticism about the fittingness of our emotions is not rational for the same reason that radical skepticism about the truth of our beliefs is not a rational position.[13]

However, even if radical emotion skepticism is not justified, a lesser form of skepticism might be rational. It needs to be admitted that many kinds of

[12] In *Epistemic Authority*, chapter 2, I argue for the rationality and inescapability of epistemic self-trust. In chapter 3, I argue that consistent epistemic self-trust commits us to have epistemic trust in others. In chapter 4 I give the parallel argument for emotional self-trust and trust in the emotions of others. I give the same argument for emotional self-trust in "Emotional Self-Trust," forthcoming, 2014.

[13] I do not actually argue that radical epistemic skepticism is irrational, but that it is less rational than epistemic self-trust. My argument depends upon my claim that rationality is doing a better job of what we naturally do anyway.

emotion are less trustworthy than many classes of beliefs because these emotions do not pass the test of conscientious reflection. Romantic love, anger, and certain forms of fear frequently do not survive reflection. Presumably, emotions like hate, envy, jealousy, and feelings of vengeance are often, and possibly always, unfitting. Lately, some writers have raised doubts about the trustworthiness of disgust.[14] Nonetheless, there are many emotions we trust, and our ability to act morally would be crippled if we were not able to place basic trust in these emotions. In this category I would place sympathy, compassion, and indignation. It is hard to imagine a functional moral life without trust in these emotions. I think that admiration is in this category. My position is that admiration is not only the basis for many moral judgments, but it is also the basis for other moral responses, particularly imitative behavior.

But this does not yet show that we ought to trust admiration, only that we are strongly inclined to trust it and often use it as the basis for the way we behave. Unfortunately, we sometimes make mistakes in admiration, even from our own point of view, as when we change our mind about who is admirable. We also know that individuals and communities disagree about the object of admiration. Some individuals admire persons whom others find dis-admirable or even contemptible. What should we say about the fact that Hitler was admired by thousands? Should those persons have trusted their emotion? Should we respond to their failure by losing trust in our own disposition to admiration?

In Claudia Koonz's interesting study of the moral conscience of the Nazis (Koonz 2003), she finds a wide correspondence between the people we (non-Nazis) admire and the people the Nazis admired. Apparently, they recognized most of the same virtues in exemplars that we recognize. The problem, in Koonz's view, is that human beings have the capacity to rule out of the moral community whole classes of persons, and many groups of people in history have exercised that capacity, leading to genocide and other heinous crimes. The biggest moral revolution of modern times is the moral attack on the exercise of that capacity, an attack that was led by moral exemplars, widely admired for qualities that even Nazis recognized as admirable. I cannot say with certainty that self-reflective Nazis could have figured out that there was something wrong with Hitler by comparing him with other persons they admired, but that is my hypothesis. There was something internally inconsistent in the set of persons they admired. In any case, there is nothing more the Nazis could have done but to conscientiously reflect upon their admiration for

[14] The research of Haidt et al. (1993) shows that people who refer to disgust as a reason not to do something cannot justify their judgment any further than the experience of the emotion of disgust itself. Martha Nussbaum (2004) has given an extended argument that disgust should be irrelevant to the law, but she clearly thinks that disgust is often an inappropriate response to a situation, quite apart from its legal ramifications. Do moral exemplars feel disgust? I think that their reactions are relevant to the trustworthiness of the emotion, and deserve further study.

Hitler, to compare Hitler with other persons they admired, and to compare their reaction to Hitler with the emotions of others they trusted. Anything they could count as evidence against the fittingness of their emotion would be discoverable through the process I call conscientious reflection.

When we consider extreme failures in admiration, that is bound to make us humble about our own disposition to admiration, but I find it interesting that when people mention the Nazi admiration for Hitler or the adolescent admiration for pop stars, nobody ever says that maybe the Nazis and adolescents are right and we are wrong. We trust our dispositions to feel admiration and contempt when they survive reflection, and I think we are right to do so. But the Nazi experience is reason to investigate the causes of errors in admiration, not just out of an unhelpful desire to blame, but because the regulation of our own emotions requires attentiveness to those causes.

When I say that a test of the justification of admiration is survival of conscientious reflection, I do not mean that we should reflect upon an emotion every time we have it, or even very often. After all, we do not always need to *show* that our emotion is justified, even to ourselves. I just mean that survival of conscientious reflection, reflection upon one's psychic states with the aim of making them fit their objects, is the test that any given psychic state is justified, and that is the test for the justification of admiration.

In other work I have argued that the trustworthiness of any instance of an emotion, including admiration, entails that I must trust the same disposition in others when it satisfies the conditions for trusting my own emotion (2012, chap. 4). To the extent that I conscientiously believe that certain other persons have the same quality of emotional conscientiousness that grounds my trust in my disposition to feel admiration, it follows from the principle that I ought to treat like cases alike that I ought to trust the emotion disposition of these other persons. This test is enough to rule out trusting the Nazi's admiration for Hitler, but if I believe that Sam is as trustworthy as I am in his disposition to feel admiration, and Sam feels admiration for Ben, I have a *prima facie* reason to admire Ben. Of course, emotion has a feeling component, and coming to believe that someone else has a given emotion is rarely sufficient to cause me to have the same emotion (leaving aside cases of empathy), but the question is whether I have a reason to judge that Ben is admirable. I think that roughly, the answer is yes. I have a *prima facie* reason to think that Sam's emotion fits its object, and so I have a *prima facie* reason to judge that Ben is admirable even if I do not feel admiration for Ben.

This result has the consequence that emotion disagreement is a problem for the conscientious person for the same reason that disagreement in belief is a problem. In my experience, we are much less likely to trust the emotions of others than to trust their beliefs, and I find that interesting. In cases in which we trust our own disposition to have a certain emotion, we do not generally extend the same trust to the emotions of other persons, even when we have

the same grounds for trusting the other as ourselves, there are no relevant differences in our circumstances, and we agree on the relevant descriptive facts. But at the same time, we often unconsciously adopt the emotions of others, particularly those we trust and admire. There is plenty of empirical evidence of the way we pick up emotions from other persons (e.g., Walden 1991; Thompson 2006). If we already do it unreflectively, it seems to me to be better to do it reflectively. Not only is admiration that survives conscientious reflection rationally justified, but reflection also reduces the chances that we will make the mistakes to which we are prone when we admire unreflectively.

6. Making Admiration Primary

I am currently developing an ethical theory based on exemplars, or paradigmatically admirable persons. Since we identify the admirable by the experience of admiration, admiration grounds the theory. I have outlined the theory in previous work[15]; here I will simply summarize the way the emotion of admiration functions in the foundation of the theory, and will indicate some advantages of a theory based on admiration.

My model for the foundational move in constructing a theory of this kind is Putnam and Kripke's theory of direct reference, which produced a revolution in semantics in the seventies. The basic idea is that a natural kind term like "water" or "gold" or "tiger" refers to whatever is the same kind of thing or stuff as some indexically identified instance. For instance, gold is whatever is the same element as *that*, water is whatever is the same liquid as *that*, a tiger is whatever is a member of the same species as *that*, and so on, where in each case the demonstrative term "that" refers directly—in the simplest case, by pointing. One of the main reasons for proposing this account of reference was that Kripke and Putnam believed that often we do not know the nature of the referent, and yet we know how to construct a definition that links up with its nature. We may not know the nature of gold—its deep structure—and for millennia nobody did, but that did not prevent people from defining "gold" in a way that fixed the reference of the term and continued to do so after it was discovered what distinguishes gold from other elements. In fact, we could not say that modern humans "discovered" the nature of gold unless we thought that modern speakers know the nature of the same stuff of which people used to be ignorant. The theory of direct reference has the advantage of explaining how "gold" referred to the same thing before and after the discovery of the molecular structure of gold.

[15] I present a Christian form of the theory in Zagzebski 2004, and outline the theory as a whole in Zagzebski 2010 and 2013, as well as in a booklength project in progress.

One of the exciting features of this proposal was that it meant that competent speakers of the language can use certain kinds of terms to successfully refer to the right things without going through a descriptive meaning. Compare a term like "hammer." When you say "hammer," you refer to whatever satisfies a description given in advance. Presumably you cannot talk about hammers unless you grasp that description. In contrast, speakers need not associate descriptions with terms like "water" and "gold" in order to successfully refer to the right kinds. In fact, they can succeed in referring to water and gold when they associate the wrong descriptions with terms like "water" and "gold." It is not even necessary that every speaker be able to identify water and gold reliably themselves as long as some speakers in the community can do so and the other speakers rely upon the judgment of the experts.

I am proposing a theory in which basic moral concepts are anchored in exemplars of moral goodness, direct reference to which are foundational in the theory. Good persons are persons *like that*, just as gold is stuff *like that*. Picking out exemplars fixes the reference of the term "good person" without the use of descriptive concepts. Good persons are identified by the emotion of admiration. If that emotion survives conscientious reflection and we find it trustworthy, then we trust the judgment "That is an admirable person." So I suggest that persons who are most admirable can comprise a non-conceptual foundation for a comprehensive ethical theory. Basic moral concepts such as *virtue, right act, good aim*, and *good life* can be defined via direct reference to exemplars, or the most admirable persons. Virtues are traits of those persons. Right acts are acts that those persons believe there is most reason to do in a given set of circumstances. The lives they live are good in the sense of admirable. The lives they desire are good in the sense of desirable. All basic moral concepts are linked to the behavior or psychology of exemplars, and the content of a virtue, a right act, a desirable life, and so on, can be discovered by investigation of exemplars.

In this theory it is not necessary for ordinary people engaged in moral practice to know the nature of good persons—what makes them good. In fact, it is not necessary that anybody knows what makes a good person good in order to successfully refer to good persons, any more than it was necessary that anybody knew what makes water, water to successfully refer to water before the advent of molecular theory. We need not associate any descriptive meaning with "good persons," and users of our language can successfully refer to good persons even when they associate the wrong descriptions with the term "good person." As with natural kinds like gold and water, people can succeed in referring to good persons as long as they, or some people in their community, can pick out exemplars. We already have practices of picking out exemplars in narratives of fictional and non-fictional persons. These practices both allow us to identify exemplars and motivate us to want to imitate them.

One of the most interesting features of the Kripkean account of natural kinds is the way empirical investigation can reveal natures, and I think this can also be a feature of exemplarist virtue theory. If the concepts in a formal ethical theory are rooted in a person, then narratives and descriptions of that person are morally revealing. It is an open question what it is about the person that makes him or her good. For the same reason, when we say that a good person is a person like that, and we directly refer to St. Francis of Assisi, or to Confucius, or to Jesus Christ, we are implicitly leaving open the question of what properties of Francis, Confucius, or Christ are essential to their goodness. I think that we believe in advance of investigation of any particular good person that what makes good persons good is related to their motivational structure, just as we think in advance of observation that what makes water water is its physical structure. In both cases we think that what we are after is not superficial properties, but deep structure. We find that out through observation. Observation tells us what the motivational structure of a person is, just as observation tells us what the molecular structure of water is.[16]

Since the feeling of admiration of persons is a kind of attraction that carries the impetus to imitate with it, a person who is admirable in some respect is imitable in that respect. That means that the ways in which exemplars are admirable, and hence imitable, can be used to give us both a way of understanding significant moral concepts, and a way of using those concepts to make ourselves and our lives conform to the admirable. Admiration has both a theoretical and a practical role in our moral lives.

To review:

1. An exemplar is a person who is paradigmatically good—most admirable.
2. We identify the admirable through the operation of the emotion of admiration.
3. A person who is admirable in some respect is imitable in that respect.
4. The features of exemplars—their traits, desires, aims, and processes of reasoning about what to do—tell us what virtues, good aims, right acts, and good lives *are*.

A theory constructed around direct reference to exemplars has a number of advantages:

---

[16] Keith Donnellan argued in unpublished work in the late seventies that we do not get Kripke's conclusion that "water is H2O" is necessary a posteriori without additional arguments than those provided by Kripke or Putnam. Nathan Salmon ([1981] 2005) discusses Donnellan's argument and says that we get the desired conclusion provided that we know a priori that what makes water water is its deep physical structure. The parallel argument can be made about the structure of a good person, but I am not defending the parallel point since the class of beliefs we acquire from trust in admiration does not fit the traditional divisions of a priori and a posteriori knowledge.

1. The foundation of the theory is not conceptual, so it does not beg any questions about the concepts of most significance to ethics. We find out from investigation of exemplars what their traits are, what they think they have most reason to do in certain situations, and what kind of life they desire. The theory does not presuppose that these substantive issues are settled in advance of the construction of the theory.
2. The perennial problem of why be moral does not arise in this theory because a motivating state generates the foundation of the theory. Admiration is both a conceptual basis for a theoretical construction and a state that moves us. As long as we admire moral persons in the way indicated in Haidt's studies, admiration moves us to want to be moral.
3. Empirical studies are directly connected with the theoretical framework. In fact, most of the substance of the theory is provided by the results of observation, both in the form of narratives that are past down from previous generations, and in the form of controlled empirical studies. The theory therefore serves well the current interest in incorporating empirical work into moral philosophy, and the interest of many moral philosophers in utilizing a narrative approach to ethics.
4. Exemplarism makes a direct connection between theoretical ethics and moral education since people naturally learn through imitation, and the theory makes the imitable the basis for the theory. In this theory there is a close connection between the process of acquiring moral concepts and the process of becoming a moral person.

My position is that admiration can be the basis for a robust and comprehensive ethical theory with the advantages I have listed, but I have not demonstrated that possibility in this brief chapter. Nonetheless, I hope that the connection between admiration and the admirable, and between the admirable and exemplars who can be carefully studied, is highly plausible, and will spur further work by moral philosophers, empirical scientists, and education theorists on admiration and its objects. I already know of some interdisciplinary work on the psychology and neuroscience of moral exemplars,[17] and work on emulation of exemplars. I hope to see more empirical work on admiration and the connection between admiration and emulation, as well as work on the distinctive features of exemplars.[18]

---

[17] See, for instance, the ongoing work by the interdisciplinary HABITVS research group consisting of Michael Spezio, Warren Brown, Kevin Reimer, James Van Slyke, and Gregory Peterson. Recently, Dirk Schumann, Steven Quartz, and Jan Glascher have joined the group. Some of their earlier work is reviewed in papers in Van Slyke, 2013.

[18] I am very grateful for comments by Christian Miller and the audience at the Wake Forest University Character Project conference in 2013. The paper was made possible through the support of a grant from The Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

References

Ackrill, J. (1980). "Aristotle on *Eudaimonia*." In *Essays on Aristotle's Ethics*, edited by A. O. Rorty, 15–33. Berkeley: University of California Press; originally published in 1974 in *Proceedings of the British Academy* 60.

Algoe, S., & Haidt, J. (2009). "Witnessing Excellence in Action: The 'Other Praising' Emotions of Elevation, Gratitude, and Admiration." *Journal of Positive Psychology* 4: 105–127.

Baehr, Jason. (2010). "The Structure of Open-Mindedness." *Canadian Journal of Philosophy* 41: 191–213.

Bilger, B. (2012). "The Strongest Man in the World." *The New Yorker*, July 23.

Colby, A., and Damon, W. (1992). *Some Do Care: Contemporary Lives of Moral Commitment*. New York: Free Press.

Chiger, K. (2008). *The Girl in the Green Sweater: A Life in Holocaust's Shadow*. New York: St. Martin's Press.

Jefferson, T. (1975). "Letter to Robert Kipwith." In *The Portable Thomas Jefferson*, edited by M. D. Peterson, 349–352. New York: Penguin; originally published in 1771.

Haidt, J. (2003). "Elevation and the Positive Psychology of Morality." In *Flourishing, Positive Psychology and the Life Well-Lived*, edited by C. L. M. Keyes and J. Haidt, 275–289. Washington, DC: American Psychological Association.

Haidt, J., Koller, S., and Dias, M. (1993). "Affect, Culture, and Morality; or, Is It Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology* 65(4): 613–628.

Koonz, C. (2003). *The Nazi Conscience*. Cambridge, MA: Harvard University Press.

Kraut, R. (1989). *Aristotle on the Human Good*. Princeton, NJ: Princeton University Press.

Marshall, R. (1990). *In the Sewers of Lvov*. New York: Harper Collins.

Nussbaum, M. (2004). *Hiding from Humanity: Disgust, Shame, and the Law*. Princeton, NJ: Princeton University Press.

Salmon, N. [1981] (2005). *Reference and Essence*. 2nd ed. Amherst, NY: Prometheus Books.

Snow, C. P. (1958). *The Search*. London: MacMillan.

Thompson, R. (2006). "The Development of the Person: Social Understanding, Relationships, Self, Conscience." In *Handbook of Child Psychology*, 24–98. 6th ed. *Vol. 3, Social, Emotional, and Personality Development*, edited by N. Eisenberg. Hoboken, NJ: Wiley and Sons.

Van Slyke, J., Peterson, G. R., Reimar, K. S., Spezio, M. L., and Brown, W.S., eds. (2013). *Theology and the Science of Moral Action: Virtue Ethics, Exemplarity, and Cognitive Neuroscience*. New York. Routledge.

Vianello, M., Gallani, E., and Haidt, J. (2010). "Elevation at Work: The Effects of Leaders' Moral Excellence," *Journal of Positive Psychology* 5, 390–411.

Walden, T. (1991). "Infant Social Referencing." In *The Development of Emotion Regulation and Dysregulation*, edited by Garber and Dodge, 69–88. Cambridge Studies in Social and Emotional Development. Cambridge: Cambridge University Press.

Zagzebski, L. (2003). "Emotion and Moral Judgment." *Philosophy and Phenomenological Research* 66: 104–124.

———. (2004). *Divine Motivation Theory*. Cambridge: Cambridge University Press.

———. (2006). "The Admirable Life and the Desirable Life." In *Values and Virtues*, 53–66, edited by Timothy Chappell. New York: Oxford University Press.

Zagzebski, L. (2010). "Exemplarist Virtue Theory." *Metaphilosophy* 41(January): 41–57. Reprinted in *Virtue and Vice: Moral and Epistemic*, 39–56, edited by H. Battaly. Oxford: Wiley-Blackwell.

———. (2012). *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. New York: Oxford University Press.

———. (2013). "Moral Exemplars in Theory and Practice." *Theory and Research in Education* 11(2): 193–206.

———. 2014. "Emotional Self-Trust." In *Emotion and Value*, New York: edited by S. Roeser and C. Todd. New York: Oxford University Press.

———. *Exemplarist Virtue Theory*. Manuscript in progress.

PART } V

Virtue Epistemology

12 }

People Listen to People Who Listen: Instilling Virtues of Deference

Kristoffer Ahlstrom-Vij

1. Introduction

When we listen to someone and believe what we are being told because they're saying it, we can be said to be *deferring* to that someone.[1] In many cases, deferring to others is a good idea. For example, when it comes to medical matters, I defer to my GP, because she knows what she is talking about. And when it comes to climate change, I defer to climate scientists, because they are best placed to answer the relevant questions.

What makes it a good idea for me to defer in these cases is that the relevant sources know what they're talking about in domains within which I'm largely uninformed. Why is that a good idea? I will follow Alvin Goldman in taking knowledge to be identical with true belief.[2] Consequently, *to know what one is talking about* is *to speak the truth*.[3] For the hearer to be *uninformed* is for her to be *in error* or *ignorant*. For example, I might be positively in error (i.e., believe something false) about the causes of climate change, while simply ignorant (i.e., not believe what's true) on the relevant medical matters prior to asking

[1] To say that someone believes what someone else is saying *because* they are saying it is to make a mere causal claim, and not to prejudge the issue of the believer's *reasons* for the relevant belief. The latter are relevant to whether or not the believer can be said to *be justified* on the basis of deference, which is not something that I will take a stand on here. However, note that, since I'll be identifying knowledge with true belief, deferring to someone who speaks the truth means coming to know.

[2] See Goldman (1999, p.5).

[3] Notice the formulation in terms of *speaking* the truth as opposed to *believing* the truth. This way of understanding what it is to know what one is talking about excludes people who *believe* the truth but are either unable or unwilling to communicate what they believe, but includes people who do *not* believe the truth but nevertheless speak the truth. As for the latter, consider Jennifer Lackey's (2008) example of a creationist teacher who teaches her students about evolution, without believing the things she's telling them.

my GP. In either scenario, when deferring to the sources in question, under the relevant circumstances, I will learn something, and thereby become (more) informed. That's why it's a good idea to defer to people who speak the truth on matters about which I'm either ignorant or in error.[4]

In fact, section 2 will suggest that being *disposed* to defer to sources who know what they're talking about in contexts where you're uninformed is not only a good idea but also an *epistemic virtue*. This is because there is a *consequentialist* notion of epistemic virtue, on which virtues simply are dispositions that promote epistemic goals, and knowledge is an epistemic goal. In section 3, it will be argued that one way to *instill* the virtue of deference is by promoting *compliance* with the requests on the part of the relevant kind of sources to be listened to. Section 4 argues that one way to promote compliance with such requests is by having those sources manifest a virtue of their own: that of *procedural justice*. Finally, section 5 responds to a number of objections.

2. Consequentialist Virtue

Let us start by getting clearer on what it is for something to be a virtue. For purposes of this chapter, I will focus on a *consequentialist* notion of virtue, on which virtues simply are dispositions that promote goals or ends. This is the notion of virtue that we find in John Stuart Mill's *Utilitarianism*.[5] While Mill thereby had a virtue *theory*, he was not a virtue *ethicist*. To qualify as a virtue ethicist, one typically has to define what makes an action right in terms of virtue. For example, a virtue ethicist might hold that I should refrain from lying simply because lying is *dishonest*,[6] or, because refraining from lying is what the virtuous person would do or recommend.[7] That's not how Mill thinks about virtue, nor is it how he thinks about right action. Since he is a consequentialist, Mill defines right action in terms of the goods generated, and the same goes for virtues, which are simply dispositions that promote moral goods or ends. As he puts it, "actions and dispositions are only virtuous because they promote another end than virtue."[8] Since Mill is a hedonist, he takes the end (or good) in question to be happiness, understood as "pleasure and the absence of pain."[9]

---

[4] Suggesting that what makes it a good idea to defer is the target of deference's knowing what she is talking about in a context where I'm uninformed might raise worries about condoning *blind deference*, i.e., deference in the absence of any insight into the epistemic credentials of the source. However, as I have defended the idea that there is nothing epistemically objectionable about such blind deference in Ahlstrom-Vij (forthcoming), I will not pursue that possibility here.

[5] See Mill (2001/1861). The relevant notion can also be found in Driver (2001).

[6] See Crisp and Slote (1997).

[7] See, e.g., Slote's (2001) agent-based virtue ethic.

[8] Mill (2001/1861: 36).

[9] Mill (2001/1861: 7).

Consider an example: let's assume that you're someone who's disposed to help people you perceive to clearly be in need. For example, when you see someone fall in the street, you'll help him or her up and make sure that he or she is alright. Provided that your perceptions on the matter are not completely off—that is, you tend to *perceive* people to be in need when and only when they actually *are* in need—that seems the virtuous thing to do. A virtue ethicist might explain why that is so with direct reference to virtue, for example by saying that it's the kind of thing that a virtuous person would do. Mill, by contrast, would explain it by saying that being disposed to help people in need in this manner is conducive to happiness. And the same would go for other prosocial dispositions, such as being generous toward others—that disposition, too, would constitute a moral virtue on Mill's theory, provided that it promotes happiness.

Mill famously thought that happiness is not just one end among many. He holds "that pleasure, and freedom from pain, are the only things desirable as ends; and that all desirable things [. . .] are desirable either for pleasure inherent in themselves or as means to the promotion of pleasure and the prevention of pain."[10] But Mill does not thereby deny that there are cases in which we value virtue in itself, irrespective of its promotion of happiness. To the contrary, utilitarians, according to Mill, "not only place virtue at the very head of the things which are good as means to the ultimate end, but they also recognize as a *psychological* fact the possibility of its being, to the individual, a good in itself, without looking to any end beyond it."[11] However, here as elsewhere, we must distinguish between *valuing* something for its own sake, and something being *valuable* for its own sake. For example, I might value money for its own sake, but it doesn't follow that money is valuable for its own sake. Arguably, money is only valuable *instrumentally*, and specifically on account of the things that it enables us to buy. And the same, Mill would argue, goes for virtue: while we might *value* it for its own sake, it doesn't follow that it's *valuable* for its own sake. To Mill, only happiness is valuable for its own sake.

Mill also is not denying that virtues can be accompanied by certain *motivations*. For example, in Aristotle, it might be a motivation to do the virtuous thing for the sake of *to kalon*—that is, the fine, noble, or beautiful.[12] Moreover, someone attracted to Mill's virtue theory does not need to deny that virtues tend to be *acquired* in certain ways (if indeed they are), for example, through experience. What Mill would deny is simply that these things are *necessary*

[10] Mill (2001/1861: 7).

[11] Mill (2001/1861: 36; my emphasis).

[12] For example, in *Nichomachean Ethics*, Aristotle says that the courageous person will "withstand [the things that make for boldness] in the way one should [. . .] for the sake of achieving what is fine; for this is what excellence aims at" (1115 b 12–13), and that "the munificent person will incur [. . .] expenditure for the sake of the fine, since this is a shared feature of the excellences" (1122 b 7–8).

for someone possessing virtue. All that is required for virtue possession is the relevant disposition promoting happiness.

So far, we have talked about *moral* virtue. But by taking Mill's theory of moral virtue as our model, we may develop a consequentialist theory of *epistemic* virtue, which we in turn may apply to dispositions to defer. Elsewhere, I have defended the idea that true belief is not only an epistemic goal but *the* epistemic goal.[13] If true belief is the epistemic goal, then provided I don't already believe what's true in the relevant domain, I will achieve that goal insofar as I defer to people who know what they're talking about, that is, who speak the truth. Moreover, if I'm positively *disposed* to defer thus—that is, if I tend to listen to people who know what they're talking about in domains where I'm uninformed—I possess a *virtue of deference*. If, by contrast, I tend to listen to people who *don't* know what they're talking about, on account of speaking falsely, I can be said to possess a *vice of deference*, and the vice of *gullibility* in particular.[14]

A word is in order about what is and is not implied by taking virtues (and vices) to be *dispositions*. Dispositions behave in the manner of *functions*. For example, if I'm disposed to believe what people tell me, then it will, under relevant circumstances, be the case that, if people tell me that *p*, I will tend to believe that *p*. In other words, testimonial input will, together with facts about my psychology, tend to generate beliefs corresponding to that input. In this case, the disposition is the psychological component of the function. Such dispositions might be stable across time, including while not subjected to any input. For example, if I am disposed to trust others, I will be so disposed even while asleep (and, as such, not receiving any testimony). That said, the temporal stability is not a necessary feature of dispositions. I might develop a severe form of paranoia that will lead me to trust no one. Later, I might be cured of my paranoia. Here, we have a case where a disposition disappears and then reappears over time.

Thinking about dispositions on the model of functions should also make clear that, even in cases where some set of dispositions remains constant over time, the corresponding virtues might not. Consider, again, my disposition to trust others. For that disposition to amount to a virtue, it would need to make me tend to attain true belief. If I happen to be more or less constantly surrounded by people telling me the truth, then it will. However, if my situation changes to a less epistemically hospitable one, then my disposition might cease to be a virtue, since it might now have me listen to people who *don't* know what they're talking about. Indeed, it might even turn into a *vice*, if it

[13] See Ahlstrom-Vij (2013).

[14] This corresponds fairly well to Fricker (2006), who suggests that someone is gullible "if she has a disposition or policy for doxastic response to testimony which fails to screen out false testimony" (p. 620).

operates in a manner that drives me deep into error or ignorance. This is to say that, while dispositions might in many cases be stable across time (although that's not a necessary feature of dispositions), and virtues are dispositions, it doesn't follow that virtues are stable over time. As we have seen, the reason is that virtues might not be stable across *environments*, and a person might find herself in different environments at different *times*.[15]

We should not think of this instability as a problem. To the contrary, insofar as we are interested in *instilling* epistemic virtues—certainly a concern of virtue theory going back all the way to Aristotle—what virtues we possess at any given moment better be something that can be made to change. On the consequentialist model mentioned earlier, such a change can be brought about either by changing the person's psychological dispositions, or by changing her environment. (In the sections that follow, I will focus on the latter, when considering how our social environment plays a role in triggering the activation of some dispositions rather than others.) Moreover, as in the case of moral virtue, accepting such a consequentialist theory of epistemic virtue doesn't require denying (a) that we, as a matter of psychological fact, might value epistemic virtues independently of that to which they are conducive; (b) that epistemic virtues might often be accompanied by certain motivations, for example, a motivation to attain true belief for its own sake; or (c) that epistemic virtues tend to be acquired in particular ways. Accepting the theory would simply involve maintaining that neither motivations nor any particular mode of acquisitions is *necessary* for possessing epistemic virtue, and to hold that valuing something for its own sake does not *necessarily* make it intrinsically valuable.

It should be stressed that defending a consequentialist notion of epistemic virtue is not to suggest that that is the only legitimate notion of epistemic virtue. There might be other, equally legitimate notions of virtue, including ones rooted in the Aristotelian tradition.[16] The only thing implied by a commitment to said consequentialist notion is that it is at the very least *among* the legitimate

[15] When on the topic of stability across environments, a word is in order on the 'situationist' challenge to virtue ethics. Very briefly, the challenge is that we have empirical reason to believe that what people do is a function of their situation to an extent that seems incompatible with maintaining that our behavior is generally a result of moral character traits, of the kind Aristotelian virtue ethicists are concerned with (Doris 2002). However, if we individuate virtues partly with reference to situations (i.e., environments), in the manner that I have done here, situational influences cease to present a challenge to the explanatory and predictive powers of virtue theory, and instead provide a practical resource for explaining behavior (doxastic and otherwise) in terms of the virtues. Moreover, since I am providing a robustly *consequentialist* virtue theory, there is no worry about some of the situational factors being 'of the wrong kind' (e.g., ambient sounds or smells); whatever makes us more disposed to achieve epistemic goods is a candidate for contributing to virtue. In that respect, my theory is potentially revisionary, as far as our common-sense conception of epistemic virtue is concerned (to the extent that we *have* such a conception).

[16] See, e.g., Zagzebski (1996); and Baehr (2011).

notions of epistemic virtue.[17] Here, as elsewhere, the proof is in the eating of the pudding. Hopefully, the sections that follow constitute a convincing case in terms of the kinds of things that we can do with such a consequentialist notion, when it comes to thinking about how we might help each other become more informed by listening to people who know what they're talking about. This is not to suggest that we cannot do these things with reference to other notions of virtue, such as Aristotelian notions.[18] It is hopefully the case that there are many things we can do, and many epistemological frameworks we can use, for purposes of generating and spreading epistemic goods. The present investigation outlines one worthwhile thing we can do, without thereby suggesting that it is the *only* thing we can do.

3. Virtue through Compliance

There are a number of questions we may ask about the virtue of deference. As already noted, the one that will be addressed here is the following: how can we *instill* that virtue? In the previous section, we noted that the possession of a virtue is a function of both the psychology of the person and her environment. In what follows, I will focus on the possibility of triggering the activation of certain dispositions by changing the person's *social* environment in particular. This focus is motivated by a problem of discrimination. Often, it is going to be hard for the individual to determine exactly *when* to defer. This problem will be particularly pressing in cases where the expertise (real or putative) of the sources lies outside of the expertise of the person considering whether to defer, such as in the case of a layperson considering what expert to consult on some matter.[19] If a person who's uninformed on the relevant matters tends to defer to sources who know what they're talking about, she instantiates the *virtue* of deference. When we ask the question of how to *instill* that virtue, and our focus is on her social environment in particular, we are putting the onus, not on the individual agent to figure out how to discriminate between good and bad sources, but on good sources to bring about deference by instilling the relevant virtue in individuals. Differently put, we are asking, what can good sources

[17] See Battaly (2012); Baehr (2011); and Greco and Turri (2011) for recent endorsements of such pluralism about legitimate notions of epistemic virtue.

[18] See, e.g., Roberts and Wood (2007).

[19] Saying that the layperson in such a scenario faces a problem of discrimination is not to deny that there are indirect strategies for gauging the epistemic credentials of experts by laymen, e.g., along the lines suggested by Goldman (2001). However, as Goldman himself acknowledges, this is by no means a trivial matter: "[T]he situations facing novices are often daunting" and raise questions about "[w]hat kinds of education, for example, could substantially improve the ability of novices to appraise expertise, and what kinds of communicational intermediaries might help make the novice-expert relationship more one of justified credence than blind trust" (p. 109).

do in order to see to it that people become disposed to defer to them on the relevant matters, and, as such, manifest the virtue of deference?

The answer provided in what follows is that we may do so through a particular form of compliance. The idea behind this answer is that what sources are doing is, in effect, providing a request together with a piece of information. Consider, for example, a professor teaching an introductory class in ethics telling her students: "Listen, all utilitarians are consequentialists, but not all consequentialists are utilitarians." When the students listen to her, they are complying with her request to be listened to. Moreover, such compliance promotes deference, given the empirical assumption that listening is conducive to believing what one is being told. This assumption is compatible with occasional failures to believe what we are being told, as long as getting us to listen still serves to increase the chances that we will believe what we are being told. So, the relevant empirical assumption is this: while we cannot be made to *believe* things, we can be made to *listen*, and trust that there is an imperfect but still robust enough connection between listening and believing the content of what's said for listening to increase the chances of belief.

It might be objected that, even if we assume that we can be brought to defer by being brought to listen, bringing us to defer to the sources we should be deferring to—that is, to those who know what they're talking about—is really only half the problem of our failure to defer in ways that we should. The other half is that of also bringing us *not* to defer to bad sources—that is, sources who *don't* know what they're talking about. Notice, however, that the two halves of the problem are not independent of one another. In particular, the more successful good sources are with respect to bringing us to defer to them, the more successful they will be with respect to bringing us *not* to defer to bad sources. This is so under the assumption that there is a practical limit to how many sources we can defer to. Given such a limit, increased success in bringing us to defer to good sources will, in effect, also promote the goal of bringing us to *only* defer to such sources. In other words, good sources can work to 'crowd out' bad sources, given practical limitations on how many sources we can defer to.[20]

If the question of how we can instill the epistemic virtue of deference can be addressed in terms of the question of how to bring about compliance with requests for good sources to be listened to, our question now becomes how to promote the relevant form of compliance. The suggestion developed in the next section is that there is a kind of *procedural justice*[21] that, if practiced by (good) sources, will achieve two things. First, it will increase the extent to

[20] We will revisit this point in section 5.

[21] See Ahlstrom-Vij (2014) for a discussion of the relation between this notion of procedural justice and other epistemologically relevant notions of justice, e.g., in Fricker (2007); Coady (2010).

which we *consult* them on the relevant matters. Second, it will increase the rate of *compliance* with their requests for us to listen to them, which will be correlated with believing what we are told, under the assumption that listening is conducive to believing. In the next section, we will look at the evidence supporting this claim, as well as at the relevant kind of justice.

4. The Virtue of Procedural Justice

As already noted, the hypothesis to be defended in the remainder of the chapter is that there is a kind of procedural justice that, if practiced by sources who know what they're talking about, will increase not only the extent to which they are consulted on the relevant matters, but also the rate of compliance with their requests for hearers to listen to them. Under the assumption that listening is conducive to believing, this will in turn be correlated with hearers' believing what they are being told. But why consider this hypothesis at all plausible? And what is the relevant notion of justice? Let us consider each question in turn.

The main source of support for the hypothesis comes from Tom Tyler and colleagues' research into the why people follow the law. On a traditional picture, people follow the law because of a fear of sanctions. However, in a landmark study, Tyler showed that fear of sanction is not the only or even the most important factor behind why people follow the law.[22] An equally if not more important factor is people considering the relevant authorities *legitimate*. To be legitimate is to be deserving of deference, and to be perceived as legitimate is to be perceived as being thus deserving. Consequently, perceived legitimacy is a property that, "when it is possessed, leads people to defer voluntarily to decisions, rules, and social arrangements."[23] More specifically, as perceived legitimacy goes up, so does compliance.[24] Moreover, the central factor determining whether people conceive of an authority as legitimate is the extent to which they take that authority to be *just*.

This brings us to the second question, regarding the content of the relevant notion of justice. The relevant kind of justice is *procedural* justice, specifically as it pertains to the following four components:

The first component pertains to being provided with an opportunity to state one's case. Let us refer to this condition as *input*. According to Tyler, "[p]eople have a tremendous desire to present their side of the story and value the opportunity in and of itself."[25] As such, being given an opportunity to state

[22] See Tyler (2006a).
[23] Tyler (2006b: 376).
[24] See Tyler (2006a: 57).
[25] Tyler (2006a: 147).

one's case has a significant positive effect on perceived justice, even when we are aware that we have no influence over the authority's actual decision on the matter (e.g., about whether we are to be fined, sentenced, etc.).[26]

The second component corresponds to the fact that, while people do not require that their input be reflected in the ultimate decision, they must be able to infer that what has been said at the very least has been considered.[27] Let us refer to this condition as *consideration*. As discussed by Tyler, the relevant condition is fairly weak: in many cases, simply explaining that the input was considered but, unfortunately, could not influence the decision is sufficient for satisfying the consideration condition.

The third component is that one's input has been evaluated in a way that is sensitive only to the facts of the matter, and not to the authority's personal preferences or prejudices. Let us refer to this as *factuality*.

The fourth component involves the relevant authority having made an effort to satisfy the above conditions—that is, *input, consideration*, and *factuality*—as opposed to simply having gone through the motions. Let us refer to this component as *effort*.

If people feel that they have been treated in a just manner—which is to say that they feel that the *input, consideration, factuality*, and *effort* conditions just outlined have been satisfied—that will lead them to consider the authority more legitimate, as well as increase their compliance with the laws laid down by that authority. In fact, perceptions of justice can give rise to a feedback loop, in that, when authorities are viewed as legitimate, their actions are more likely to be seen to be just.[28, 29]

The following are two extensions of these results that are relevant presently:

First, the relationship between perceived justice, perceived legitimacy, and compliance is not unique to the legal domain; it applies to *rule following* more generally. For example, employees' perceptions regarding the justice of corporate policies, and the resulting perceptions of the legitimacy of their employers, track employees' compliance with corporate policies,[30] as well as cooperative behavior more generally, including compliance with rules in the absence of explicit

---

[26] Tyler (2006a: 127).

[27] See Tyler (2006*a*: 149).

[28] See Tyler (2006a: 107).

[29] To prevent confusion for readers familiar with Tyler's work, it should be noted that in some places (e.g., in Blader and Tyler 2003), Tyler talks about a four-*component* model of procedural justice. In those contexts, the components correspond to the *dimensions* along which people evaluate procedural justice—namely, with respect to decision-making, quality of treatment, formal interactions, and informal interactions. By contrast, the four components discussed here pertain to what Tyler and colleagues' research suggests constitutes the *content* of justice evaluations, in the sense of what we look for when evaluating people's behavior or the structure of organizations, along aforementioned dimensions or otherwise.

[30] See Tyler and Blader (2005).

requests for compliance or threats of sanctions.[31] The same goes for non-corporate employees, such as police officers and members of the military, who, too, are significantly more inclined to comply with attempts to regulate their performance insofar as they deem their employers just, and thereby legitimate.[32]

Second, perceived legitimacy is not only relevant to our tendency to comply with communicated rules or norms—that is, doing what we are being told to do—but also to our willingness to engage in *consultation*, that is, seeking advice on what is to be done in the first place. For example, on the question of *whom* to consult, people report being more prone to consult professionals regarding retirement saving and investment strategies who they perceive to be just than professionals who they do not perceive to be just, even when aware that the cost of receiving a just treatment would be a decreased likelihood of financial gain.[33] Moreover, as for the question of *whether* to consult someone, students report being more likely to seek advice from their professors, on academic as well as on personal matters, when they take the professors to be such that they would treat them in a just manner.[34]

Consequently, if we were to formulate a slogan capturing what sources need to do in order to promote deference—assuming, as above, that listening is to a robust enough extent conducive to believing what's said—it would be this: *people listen to people who listen*. In formulating the slogan thus, I am making an assumption about the relationship between *perceived* justice and *de facto* justice, to the effect that the best way to *seem* just is to *be* the things that people consider just. The rationale for that assumption is that, while it is certainly conceptually possible to solve compliance problems by simply *seeming* to be those things, it is not only unlikely that one would be able to pull off the level of deception required in the long run, but also unclear what would be gained from it. After all, maintaining the relevant deception would most likely require a greater amount of resources than would simply *being* the things people think of as just. In other words, the best way to have people perceive good sources as just is for those sources to sincerely work to provide a forum for input, and to take into account what is being said in a factual manner, where the relevant

[31] See Tyler (2011: chap. 3). To avoid confusion, I have ignored a possible terminological shift between Tyler's earlier (e.g., his 2006a) and his more recent work (e.g., his 2011). In the former, what I have referred to as the *input* and *consideration* conditions are conditions on perceived procedural justice; in the latter, these are occasionally (e.g., 2011: 106) but not invariably (e.g., 2011: 114) treated as conditions on *perceived trustworthiness*, where the latter is treated as a separate contributory factor to perceived legitimacy, in addition to that of perceived procedural justice (see, e.g., his 2011: 112). Since this terminological shift makes no substantive difference, I will stick to his earlier terminology, and take perceived trustworthiness—if that is the term we should use for what is captured by *input* and *consideration*—to be a component of perceived procedural justice.

[32] See Tyler, Callahan, and Frost (2007).

[33] Tyler (2006c: studies 1 and 2).

[34] Tyler (2006c: study 3).

consideration involves also explaining why the input does not change the verdict in cases where it does not.

By way of illustration, return to the case of the professor telling her students: "Listen, all utilitarians are consequentialists, but not all consequentialists are utilitarians." The request signals that the professor wants the students to listen to her—that is, to comply with her request to be listened to—and the reason she wants that is because listening is conducive to believing, in accordance with the empirical assumption discussed earlier. So, how can she bring the students to listen to her? If what has been argued in this section is correct, she can do so by listening to her students in turn, in such a manner that they feel that they've been provided with a forum for input wherein their viewpoints are evaluated in a factual manner, and that it's being explained in what respects they are correct or mistaken. As such, the present framework confirms what most people who have been involved in education probably already know: the best way to inform people is not to tell them what to think, but to engage them in a dialogue.

We're now in a position to formulate the upshot of the present investigation. We started out by noting that there is a consequentialist notion of epistemic virtue. On such a notion, epistemic virtues are dispositions that promote the epistemic goal of believing truly. It was noted that this is not the only notion of epistemic virtue on offer—in addition to the consequentialist notion of virtue utilized in this investigation, there are also notions of epistemic virtue rooted in the Aristotelian tradition—but that it might be fruitfully applied in contexts concerned with bringing about epistemic improvements in testimonial interactions. What has been argued here has hopefully made good on that claim, by showing that there are (at least) two consequentialist virtues that can be involved in such contexts:

The first virtue is a *virtue of deference*, manifested when a hearer is disposed to defer to sources who know what they're talking about within domains where the hearer is uninformed. To return to the example used at several points above, the student who is disposed to listen to her professor manifests a virtue of deference, if the student is uninformed and her professor knows what she's talking about in the relevant contexts. By contrast, if the student is disposed to listen to people who don't know what they're talking about, on account of speaking falsely, she can be said to manifest the vice of *gullibility*. I will have more to say about this vice in section 5.

The second virtue is a *virtue of procedural justice*, manifested by informed speakers insofar as they are disposed to interact with uninformed hearers in a way that involves listening in a manner satisfying the *input, factuality, consideration*, and *effort* conditions outlined earlier. Specifically, being disposed to interact with hearers in this particular way is a virtue insofar as it increases consultation and compliance rates, and, in so doing, instills a virtue of deference in uninformed hearers. Here, it's also worthwhile to highlight the possibility of a number of

contrasting vices, corresponding to the set of dispositions that would not only serve to decrease consultation and compliance rates, but also to positively drive people into the hands of people who don't know what they're talking about. That said, I will not elaborate on the variety of vices that can be imagined here.

One final point, before moving on to a couple of objections and replies. I have been working with a consequentialist framework throughout this paper. The reader might wonder how that framework fits with the deontological-sounding vocabulary that has been introduced in this section, relying on terms like 'procedural justice'.[35] The answer is as follows: the psychological research we've been considering suggests that people's *judgments* about what's procedurally just can be captured by the *input, factuality, consideration*, and *effort* conditions. This doesn't give us a *theory* of justice, as in: a normative theory of what justice *consists* in. Since it doesn't give us a theory of justice, it also doesn't give us a *deontological* theory of justice—or, indeed, any other kind of theory of justice. It simply gives us a psychological story about how people *think* about justice, and what they *perceive* to be just as a result. When, in the present context, talking about people being procedurally just, that's all we're talking about: being whatever the relevant empirical research suggests that people take procedural justice to consist in.

The normative significance of the relevant empirical results enters when we add to that story what happens when people are perceived to be just in the relevant sense: other people are more likely to consult them and to comply with any requests they might provide. That's normatively significant, and for purely consequentialist reasons: insofar as those consulted and complied with know what they're talking about, we can expect epistemic benefits when those doing the consulting and complying are uninformed within the relevant domains. Moreover, insofar as the relevant behaviors become ingrained enough for it to make sense to talk about the former as being *disposed* to act in ways that have the sources be perceived to be procedurally just, and others as being *disposed* to consult and comply as a result, we can describe the relevant tendencies as amounting to epistemic virtues, in the consequentialist sense introduced at the outset of our investigation. That is how to square the overarching consequentialist framework with the deontological-sounding vocabulary of this section.

5. Objections and Replies

Let us consider some objections to what has been argued so far. In the previous section, it was suggested that a hearer disposed to act in the manner captured by the *input, factuality, consideration*, and *effort* conditions can under

[35] Thanks to Christian Miller for raising this question.

certain circumstances be said to be instantiating an epistemic virtue. This is so because there is a consequentialist notion of epistemic virtue on which such virtues simply are dispositions promoting the attainment of the epistemic goal of believing truly. Bringing people who are uninformed within the relevant domains to listen to sources who know what they're talking about—that is, who speak the truth—promotes that goal. Hence, being disposed to bring people who are uninformed to listen to sources who know what they're talking about by exercising the relevant kind of justice qualifies as an epistemic virtue, and a virtue of procedural justice in particular.

But here is an objection: The relevant virtue does not necessarily promote any goal for the *hearer*; in particular, it does not necessarily make *the hearer* epistemically better off. And if it does not, how can it be right to say that the relevant disposition qualifies as an epistemic virtue on *her* part?

It is true that the relevant disposition to be procedurally just does not necessarily make the hearer possessing that disposition epistemically better off. However, it is not clear why that would disqualify the disposition as a virtue. Many non-epistemic virtues are virtues primarily, if not exclusively, on account of the good that they bring to others. Think, for example, of courage and generosity. These qualify as virtues because they are beneficial to others, whether or not they are beneficial to the virtuous person herself. Being courageous and generous might certainly be beneficial for the courageous and generous individual in the long run. Courage might bring you fame and respect, while generosity might motivate people to help you out, were you ever to find yourself in a difficult situation in turn. But it is clear that the relevant virtues would remain virtues, even in the absence of any long-term benefits to the courageous or generous person. And the same goes for the relevant notion of procedural justice as virtue: it being possible that that virtue only promotes the epistemic good of others in no way disqualifies it as an epistemic virtue.

It might be objected that there are many cases in which we consult and defer to others without justice being a factor. Consider, for example, cases wherein people simply accept what prominent scientists are telling them. Notice, however, that what I have argued merely outlines a tool for *promoting* deference, which is compatible with our deferring in other cases wherein people are not utilizing that tool. The case of science is still interesting, though. It might be that, when people trust scientists, they trust science, and part of what it is to trust science is to trust that one's viewpoint (as in: its content) has been taken into account. In some cases, people might acknowledge that this condition is satisfied for fairly uninteresting reasons, as when I trust that my *nonexistent* viewpoint has been taken into account by those working on quantum mechanics or the big bang theory.[36] So the relevant trust can be captured in the following manner: to the extent that I take

[36] Thanks to Christian Miller for suggesting this qualification.

myself to have anything to contribute (and I might not), I trust that my contribution (whether actually put forth by me, or—which is more likely—by someone in the scientific community) has been taken into account. After all, it is interesting to note that what seems to be lacking in controversial cases of distrust in science (e.g., climate change deniers and creationists) is exactly trust of this kind: those who refuse to trust science seem to think that their viewpoint has *not* been taken into account.[37]

Let us consider a different worry. In section 4, I suggested that we may formulate the upshot of the present chapter in terms of the slogan that *people listen to people who listen*, if we assume that *perceived* justice and *de facto* justice are related in such a way that the best way to seem just is to be whatever it is people consider just. However, this might raise worries about bad sources—that is, sources who don't know what they're talking about—being able to come across as just and, as such, bring about consultation and deference. This worry speaks to a point made earlier. The focus of the chapter has been on how to enable good sources to bring about deference. However, as we noted in section 3, if we assume that there is a practical limit to how many sources people can defer to, then increased success in bringing people to defer to good sources will, in effect, also serve the goal of bringing people to defer *only* to good sources. We can put this point in terms of the virtues and vices of deference introduced at the end of section 4: Since it is practically impossible for people to defer to everyone, the more successful we are when it comes to instilling the virtue of *deference*—that is, a disposition to listen to good sources within domains in which one is uninformed—the fewer opportunities there will be for people to manifest the vice of *gullibility*—that is, the disposition to listen to sources who don't know what they're talking about. In other words, promoting the relevant virtue and obstructing the relevant vice are not two separate projects: to the contrary, doing the former will, in effect, contribute to the latter.

Let us consider one final objection. If what was argued here is correct, sources manifesting the virtue of procedural justice increase the chances that people will consult as well as listen to them. However, it might be objected that there is something questionable, maybe even objectionable, about that virtue. Often, we listen to people because we are interested in hearing what they are saying, and not primarily for the purpose of having those people in turn listen to us. If a person who knows what she's talking about listens to us primarily—perhaps even simply—for the purpose of having us listen to her, is there a sense in which she is *using* us?[38] I am inclined to say 'no', the

[37] Thanks to Walter Sinnott-Armstrong for raising the question, and to John Turri for the suggested link between trust and deference in scientific contexts.

[38] I'm grateful to Stephen Grimm and Rebecca Stangl for raising versions of this worry.

reason being this: while it would make sense to talk about someone using us if the purpose of getting us to listen would be for *her* to gain something (think advertisements), the contexts that concern us here involve attempts to get us to listen for the purpose of benefiting *us*. If we can be brought to listen to good sources, we are not being used—on the contrary, we are being helped. Specifically, we are being helped to become more epistemically virtuous.

6. Conclusion

We often fail to defer to people who know what they're talking about. When doing so consistently, we fail to possess a *virtue of deference*, a disposition to defer to sources who know what they're talking about. This is because, on a consequentialist notion of epistemic virtue, epistemic virtues are dispositions that promote the attainment of the epistemic goal, and the epistemic goal is knowledge. Following Goldman, I identified knowledge with true belief.

The present chapter made two points about how to go about instilling the virtue of deference. First, it was argued that one way to instill that virtue is by promoting compliance with requests on the part of good sources to be listened to, under the assumption that listening is conducive to believing. Second, it was suggested that research in social psychology on the relationship between justice and compliance suggests that one way to promote compliance with such requests is by having the relevant sources manifest a virtue of their own—namely, that of *procedural justice*, which involves as a crucial component being prepared to listen.

In closing, I would like to make two suggestions for potential, practical applications:

One area within which failures to defer have proved particularly problematic is that involving the provision of intellectual advice. In particular, while we know that we are fallible creatures, susceptible to a number of cognitive biases, we also have a strong tendency to underestimate our susceptibility to such bias, and therefore fail to defer to people providing advice on how to avoid bias. Since bias renders us susceptible to false belief, it also prevents us from attaining virtue within the relevant domains. As I have argued elsewhere, this is a problem of deference, and of *intellectual* deference in particular, that might be addressed with reference to the exercise of procedural justice.[39]

Another potential area of application concerns the dissemination of scientific information. Failures to defer to scientific experts are widespread in certain quarters, and present not merely an epistemic problem insofar as the relevant failures stem from epistemic vices promoting ignorance and error, but

[39] See Ahlstrom-Vij (2014).

moral problems as well. Consider, for example, how the failure on the part of a subset of the population to defer to climate scientists enables politicians to do the same, with the consequence that future generations are likely to suffer gravely. Here, too, it's possible that we can put to work the notion of procedural justice in order to increase popular deference, and thereby bring about political accountability, which in turn will (hopefully) lead to the taking of necessary action.[40]

References

Ahlstrom-Vij, K. (2013), "In Defense of Veritistic Value Monism." *Pacific Philosophical Quarterly* 94(1): 19–40.

Ahlstrom-Vij, K. (2014). "Procedural Justice and the Problem of Intellectual Deference." *Episteme* 11(4): 423–442.

Ahlstrom-Vij, K. (Forthcoming). "The Social Virtue of Blind Deference." *Philosophy and Phenomenological Research*.

Aristotle (2002). *Nichomachean Ethics*. Translated by C. Rowe. Edited by S. Broadie and C. Rowe. Oxford: Oxford University Press.

Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.

Battaly, H. (2012). "Virtue Epistemology." In J. Greco and J. Turri (Eds.), *Virtue Epistemology: Contemporary Readings* (pp. 3–32). Cambridge, MA: MIT Press.

Blader, S. L., and Tyler, T. R. (2003). "A Four-Component Model of Procedural Justice: Defining the Meaning of a 'Fair' Process." *Personality and Social Psychology Bulletin* 29(6): 747–758.

Coady, D. (2010). "Two Concepts of Epistemic Injustice." *Episteme* 7(2): 101–113.

Crisp, R., and Slote, M. (1997). *Virtue Ethics*. Oxford: Oxford University Press.

Doris, J. (2002). *Lack of Character*. Cambridge: Cambridge University Press.

Driver, J. (2001). *Uneasy Virtue*. Cambridge: Cambridge University Press.

Fricker, E. (2006). "Varieties of Anti-Reductionism about Testimony: A Reply to Goldberg and Henderson." *Philosophy and Phenomenological Research* 72: 618–628.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Goldman, A. (1999), *Knowledge in a Social World*, Oxford: Oxford University Press.

Goldman, A. (2001), 'Experts: Which Ones Should You Trust?', *Philosophy and Phenomenological Research* 63(1): 85-110.

[40] For helpful discussion, I am grateful to Sandy Goldberg and Baron Reed, as well as to the participants at the Final Research Colloquia of the Character Project at Wake Forest University, and in particular to Anne Baril, Stephen Grimm, Nathan King, Walter Sinnott-Armstrong, Rebecca Stangl, and John Turri. I am also grateful to Christian Miller for detailed comments on an earlier draft of this paper. This publication was made possible through the support of a grant from The Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Greco, J., and Turri, J. (2011). "Virtue Epistemology." In *Stanford Encyclopedia of Philosophy*. Available at http://plato.stanford.edu/entries/epistemology-virtue/.

Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

Mill, J. S. (2001). *Utilitarianism* (2nd ed.). Indianapolis, IN: Hackett; originally published in 1861.

Roberts, R. C., and Wood, J. (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.

Slote, M. (2001). *Morals from Motives*. Oxford: Oxford University Press.

Tyler, T. R. (2006a). *Why People Obey the Law*. 2nd ed. Princeton, NJ: Princeton University Press.

Tyler, T. R. (2006b). "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57: 375–400.

Tyler, T. R. (2006c) "Process Utility and Help Seeking: What Do People Want from Experts?" *Journal of Economic Psychology* 27: 360–376.

Tyler, T. R. (2011). *Why People Cooperate: The Role of Social Motivations*. Princeton, NJ: Princeton University Press.

Tyler, T. R. and Blader, S. L. (2005) "Can Businesses Effectively Regulate Employee Conduct? The Antecedents of Rule Following in Work Settings." *Academy of Management Journal* 48 (6): 1143–1158.

Tyler, T. R., Callahan, P., and Frost. J. (2007) "Armed and Dangerous (?): Motivating Rule Adherence among Agents of Social Control." *Law and Society Review* 41: 457–492.

Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

13 }

"Why Can't We Be Friends?"

REFLECTIONS ON EMPIRICAL PSYCHOLOGY AND VIRTUE EPISTEMOLOGY

Nathan L. King

I. Introduction

According to standard versions of *virtue ethics*, a person's moral character is the primary locus for ethical evaluation. It is also a predictor of her actions: courageous persons are likely to act courageously across a range of situations; compassionate persons are likely to act compassionately under many different circumstances; and so on. Further, the relevant character traits figure prominently into explanations of human actions: the courageous person acts courageously *because* she is courageous. Thus, virtue ethics seems committed to the claim that human agents have robust character traits such as courage, compassion, temperance, modesty, and wisdom—and that these traits explain a significant proportion of human behavior.[1]

Recently, moral philosophers have appealed to findings in social psychology to evaluate such empirical commitments. According to a prominent line of thought, there is little evidence for, and much evidence against, the claim that many human agents possess the classical moral virtues. Instead of resulting from robust moral character traits, morally relevant human behavior is highly susceptible to, and better explained in terms of, certain features of the agent's situation. Call this view *ethical situationism*.[2] If true, this sort of situationism seems to threaten virtue ethics by threatening the descriptive psychology the latter presupposes. For if situationism is true, virtue ethics appears to provide an incorrect explanation for human behavior.

[1] For helpful discussion on this point, see Miller (2014a), chapter 8.

[2] See, e.g., Doris (1998, 2002); and Harman (1999) for seminal defenses of ethical situationism. In this connection, note that the version articulated here is one among several ways of formulating the view. For a review of alternative formulations, see Miller (2014a), chapter 4.

Discussion of ethical situationism comprises a large literature.[3] But ethical theory isn't the only area of philosophy in which the notions of character and virtue are prominent. Since at least 1980, epistemologists have sought to develop *virtue epistemology* (VE)—an approach to epistemology that gives the notion of an intellectual virtue a fundamental role.[4] VE has matured into an important subfield of the theory of knowledge, encompassing differing conceptions of intellectual virtue. As we'll see, some versions of VE propose novel solutions to traditional epistemological problems. Other versions promise to lead epistemology in new directions.

Given VE's development, it was natural to expect that theorists would explore and test its empirical commitments against the findings of empirical psychology. Such an expectation has been met in a fast-growing literature concerning *epistemic situationism*, the view that human cognition typically results not from intellectual virtue or vice, but rather, is highly susceptible to epistemically irrelevant factors in a subject's environment; moreover, the intellectual character traits and faculties most human thinkers have are not the virtues that figure prominently in virtue epistemologies. If few human thinkers have or exercise intellectual virtues, one might think, this spells trouble for VE.

This chapter explores the situationist challenge to VE. In section II, I distinguish between three varieties of epistemic situationism and five varieties of virtue epistemology. These distinctions foster the explication and evaluation of arguments discussed in later sections. In section III, I explain two important situationist arguments against VE. These arguments exhibit the grounds on which virtue epistemologies of various sorts have been charged with (i) skepticism, (ii) empirical inadequacy, and (iii) normative paucity. In section IV, I argue that once the varieties of situationism and VE are properly understood, the apparent tension between the two largely dissipates. The upshot of the chapter, then, is that many projects and positions within virtue epistemology can be reconciled with the findings of empirical psychology. In section V, I propose that this reconciliation might evolve into a mutually beneficial friendship.

II. The Views

A. VIRTUE EPISTEMOLOGY

VE comes in several varieties. It is useful to sort them in two ways: first, according to their corresponding conceptions of intellectual virtue; and second, according to the use to which they put the concept of an intellectual virtue.

---

[3] See Miller (2014a) for a helpful review of this literature.

[4] The literature on virtue epistemology was inaugurated by Ernest Sosa (1980).

Theorists commonly divide the territory into *reliabilist* and *responsibilist* conceptions of intellectual virtue.[5] *Virtue reliabilists* understand intellectual virtues as stable, reliable cognitive faculties, such as accurate vision, good memory, intuition, and conditionally reliable inferential capacities.[6] In contrast, *virtue responsibilists* model the intellectual virtues on the moral virtues; virtues are stable traits of *character*. Among the intellectual virtues, responsibilists include such traits as intellectual courage, open-mindedness, intellectual humility, intellectual carefulness, intellectual charity, and the like. Such virtues are stable, excellent traits of cognitive character involving a motivation for epistemic goods (e.g., knowledge and true belief) which typically require an agent's efforts for their acquisition and maintenance.[7]

Despite these differing conceptions of intellectual virtue, it would be a mistake to present reliabilist and responsibilist VE as altogether opposed to one another.[8] But for the purpose of mapping the conceptual territory at the highest level of taxonomy, we may divide virtue epistemologies as shown in figure 13.1:

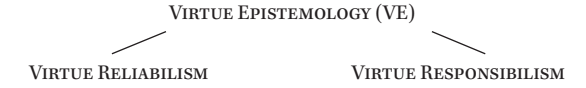

FIGURE 13.1 Varieties of Virtue Epistemology.

Below this level, the most salient division among versions of VE concerns the work that the concept of an intellectual virtue does in the theory.[9] Virtue reliabilists typically employ the concept in analyses of epistemically important properties, such as *knowledge* and *justified belief*, and in responding to skepticism.[10] Virtue responsibilists, on the other hand, put their concept of an intellectual virtue to more varied use. Some think responsibilist virtues are needed to execute traditional epistemological projects like those noted earlier. Call this position *strong conservative virtue responsibilism*—"conservative" indicating that the view sticks to traditional epistemological problems, and "strong" indicating the conviction that responsibilist virtues are essential to

[5] See Greco and Turri (2011).

[6] A *conditionally reliable* faculty or process is one that reliably yields true output beliefs, given true input beliefs. See Goldman (1979) for further discussion.

[7] For similar accounts, see Zagzebski (1996); Roberts and Wood (2007); and Baehr (2011).

[8] Arguably, a complete virtue epistemology will appeal to both reliabilist and responsibilist virtues. For one thing, it is difficult to see how responsibilist virtues could consistently lead to the truth unless the cognitive agents possessing them also possessed reliable faculties. And plausibly, at least under certain conditions, cognitive agents who possess responsibilist virtues will be better able to successfully employ their reliable faculties in order to find the truth and avoid error. See Greco and Turri (2011) and Baehr (2011).

[9] The taxonomy developed here follows that of Baehr (2011). See his helpful discussion in chapter 1 of that work, and in Baehr (2008).

[10] See, e.g., Sosa (2007, 2009); and Greco (2000, 2010).

solving these problems. Linda Zagzebski, for instance, analyzes both knowledge and justified belief in terms of responsibilist virtue. On her view, knowledge is a state of true belief arising from acts of intellectual virtue; and a justified belief is one that a person motivated by intellectual virtue, and possessing the understanding of his cognitive situation an intellectually virtuous person would have, might believe in like conditions.[11] By focusing on responsibilist virtues, Zagzebski argues, we can develop accounts of knowledge and justified belief that avoid many of the problems attending other views.[12]

Many virtue responsibilists reject this sort of research program. Such responsibilists doubt that their notion of an intellectual virtue will solve traditional epistemological problems.[13] For these theorists, responsibilist virtues enter the discussion either as *supplements* to the discussion of traditional epistemology, or as *independent* of traditional epistemology.

Weak conservative virtue responsibilists retain an interest in linking their view to traditional epistemological questions (thus, "conservative") while rejecting the notion that responsibilist virtues foster answers to such questions (thus, "weak"). There are several possible projects and positions available to such responsibilists. To mention just two: Jason Baehr has argued that responsibilist virtues can play an important but secondary role in the complete understanding of reliabilist and evidentialist theories of knowledge.[14] And Christopher Hookway has argued that the virtue of *wisdom* can help agents discern a point beyond which one ought not to take skeptical questioning seriously.[15]

Autonomous virtue responsibilism is the view that responsibilist virtues are best employed in the service of projects that are *independent* of traditional epistemological questions. Instead, such virtues are better suited to address non-traditional questions that complement or replace the questions of traditional epistemology. Like conservative responsibilism, autonomous responsibilism comes in strong and weak varieties. *Weak autonomous responsibilism* seeks to complement traditional epistemology by analyzing individual intellectual virtues (cf. Roberts and Wood 2007) and shedding light on epistemological issues that traditional epistemology has left to the side (e.g., the notion of a virtuous *inquiry* explored in Hookway 2003). *Strong autonomous responsibilism* goes further. Instead of merely seeking to complement traditional epistemology, its proponents seek to *replace* much of traditional epistemology

---

[11] On these definitions, see Zagzebski (1996) and sections III and IV.

[12] Among the issues Zagzebski takes her version of VE to illuminate: the analysis of knowledge, the analysis of justified belief, the Gettier problem, the internalism-externalism debate, and the proper response to skepticism.

[13] See, e.g., Baehr's criticisms of Zagzebski's account of knowledge in Baehr (2011, chap. 3).

[14] See Baehr (2011), chapters 4 and 5.

[15] Hookway (2003).

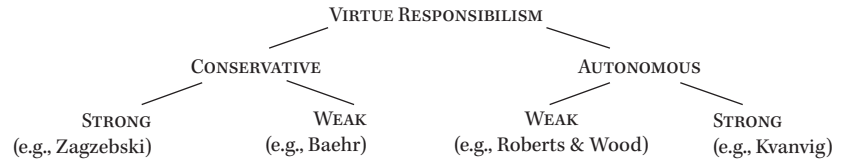

FIGURE 13.2 Varieties of Virtue Responsibilism.

with a virtue-centered picture that moves beyond the time-slice evaluations of belief that characterize the former.[16]

Virtue responsibilism, then, divides into the categories shown in figure 13.2.

These categories are helpful not only as means to conceptual understanding, but also as tools for evaluating situationist arguments. For as we'll see, these species of VE are not alike with respect to their empirical commitments, and therefore not alike in their vulnerability to the findings of empirical psychology.

B. EPISTEMIC SITUATIONISM

Like VE, epistemic situationism comes in several varieties. Some of these are uncontroversial; others are disputed. *Weak epistemic situationism* is the thesis that subtle (and seemingly epistemically irrelevant) factors in a subject's situation can influence that subject's cognitive behavior. *Strong epistemic situationism about intellectual character* (Strong Situationism-$_{CHAR}$) claims that "Most people's conative intellectual traits are not virtues" of the sort responsibilists discuss.[17] S*trong situationism about reliability* (Strong Situationism-$_{REL}$) claims that human beings lack stably reliable faculties of the sort that virtue reliabilists require for knowledge. Weak epistemic situationism is relatively uncontroversial, and enjoys a wealth of empirical support.[18] The strong versions of situationism are more controversial, as is the extent to which they are grounded in empirical psychology.

Recently, Mark Alfano has argued for Strong Situationism-$_{CHAR}$. Alfano, Lauren Olin, and John Doris have argued for Strong Situationism-$_{REL}$—all by appeal to numerous studies. And unlike Weak Epistemic Situationism, the strong versions pose at least a *prima facie* threat to VE. Accordingly, we'll now consider these views and their corresponding empirical support.

1. Strong Situationism-$_{CHAR}$

Alfano's arguments for Strong Situationism-$_{CHAR}$ examine empirical research he takes to suggest that most people lack responsibilist virtues such as curiosity,

[16] See Kvanvig (1992).

[17] Alfano (2012), p. 234.

[18] See the studies cited in Alfano (2012, 2013); and Olin and Doris 2014.

creativity, mental flexibility and courage. Space here permits discussion of only a few relevant studies.[19]

Alfano cites research by Alice Isen and colleagues as showing that many subjects fail to exhibit creativity and mental flexibility. In studies designed to test for creativity in certain problem-solving tasks, many or most subjects in the control condition fail in their tasks—thus failing to exhibit creativity. By contrast, subjects in the "positive affect condition" fare much better at the tasks after eating candy or watching a short comedy video. Thus, when such subjects behave in characteristically creative ways, their conduct is best explained in terms of their exhibiting the local traits of *flexibility while in a good mood*, or *creativity while in a good mood.* This casts doubt on behavioral explanations that appeal to responsibilist virtues, which are cross-situational and consistent.

Consider the research Isen and her colleagues conducted on the Duncker candle task.[20] Subjects are presented with a book of matches, a box full of thumbtacks, and a candle. They are then asked to affix the candle to a cork board so that when the candle is lit, no wax drips on the floor. The only solution is to empty the box and use the tacks to pin the box to the board, using the box as a shelf. Isen and colleagues found that in the control condition, only 13% of subjects solved the task. By contrast, 75% of subjects who were given candy or who watched a brief comedy prior to the trial solved the problem.[21]

Consider next the empirical research concerning intellectual courage. Alfano cites numerous studies on his way to the conclusion that most people lack intellectual courage as a global virtue (a stable, excellent disposition that reliably manifests itself under the appropriate conditions across a wide range of situations). These studies reveal that "seemingly trivial and epistemically irrelevant situational factors influence [courage relevant] epistemic and doxastic conduct" (Alfano 2012, p. 240). Alfano thinks this suggests that most people have (at best) *courage in the face of non unanimous dissent*. Central to Alfano's treatment of intellectual courage is the kind of courage needed to speak one's mind in the face of dissent. Several relevant studies seek to measure the effects of social pressure on subjects' willingness to report on the deliverances of their senses. For example, Mufazer Sherif has tested the effects of social pressure in experiments involving the "autokinetic effect."[22] Subjects are placed in a dark room in which there is a single point of light on the wall. Though the light remains stationary, when subjects' eyes wander, the light appears to move, though it's difficult to say by how much. Sherif placed subjects and confederates together in such a room. Confederates confidently affirmed an exact amount of light movement (say, three inches) and subjects always came to agree with

[19] Some paragraphs in this section are from King (2014a).

[20] Isen, Daubman, and Nowicki (1987).

[21] For discussion of further studies on creativity, see Alfano (2012).

[22] Sherif (1937).

this unanimous answer, whatever it was. Sherif infers that this social pressure affected subjects' perception of reality. Alfano opts for the weaker conclusion that "apparent unanimity can generate consent when the object of judgment is highly ambiguous" (Alfano 2012, p. 243). In other words, apparent unanimity can alter a subject's verbal behavior under the specified conditions, even if it does not alter her *perception.*

As Alfano notes, one might doubt that it's a failure of courage to account for others' opinions in making a judgment about an ambiguous matter. What about cases in which the object of judgment *isn't* ambiguous? In such cases, perhaps subjects will remain steadfast even in the face of unanimous dissent. Seeking to test this hypothesis, Solomon Asch designed an experiment in which seven confederates and one subject judged (in order and aloud) which of two lines was longer.[23] Though the correct answer was always clear, in some cases all of the confederates "judged" the shorter line to be longer. As a result, many subjects concurred with the majority, despite expressing reservations about doing so. About a third accorded with the majority more often than not, and over the course of a typical trial, between 50% and 80% of subjects went along with the majority at least once. In follow-up studies, Asch found that this effect was absent in experiments involving just one confederate and a subject, and was weak when there were two confederates and one subject. With a group of three confederates, the social effect of majority opinion was very strong, though when just one confederate went against the group, more than 90% of subjects were willing to go against a majority of confederates. Alfano doubts that the relevant subjects possess intellectual courage as a global trait. Their actions, he thinks, reveal that they have only something like *courage in the face of a non unanimous majority*—a trait that falls short of the excellence that virtue responsibilists extol. Alfano concludes that if similar results can be replicated with respect to other virtues, then human subjects lack the intellectual virtues that figure prominently in virtue responsibilist theorizing. Even when they succeed in their cognitive tasks, their behavior is often best explained in terms of epistemically irrelevant but causally efficacious factors that comprise their epistemic situation, in conjunction with the relevant local traits.

2. Strong Situationism-$_{REL}$

In their arguments for Strong Situationism-$_{REL}$, Olin and Doris appeal to studies that appear to undermine the reliability of human thinkers across a range of cognitive contexts. But because such reliability is the central feature of reliabilist accounts of intellectual virtues (and is also required by Zagzebski's

[23] See Asch (1951, 1952, 1955, 1956).

account of responsibilist virtues[24]) Strong Situationism-$_{REL}$ poses a skeptical threat to these views.

As above, presentation of the relevant empirical studies must be brief. Consider the following conclusions, cited in Olin and Doris (2014), all of which seem to show that human cognitive function varies dramatically according to context, and often does so due to epistemically irrelevant factors:[25]

- Subjects are more likely to say that a written statement is true when the color in which it is printed makes it easier to read. They are more likely to say it is false if its color makes it difficult to read.[26]
- Subjects are more likely to judge a speaker to be credible (and to agree with the speaker's arguments) if he or she speaks quickly. The phenomenon holds independently of the strength of the speaker's arguments.[27]
- Subjects predict better performance for stocks with easy-to-pronounce names than for stocks whose names are harder to pronounce.[28]
- Subjects who are tired, or who are wearing a heavy backpack, are more likely to overestimate the slope of a hill, or the distance to a specified target object.[29]
- Subjects provide consistent, accurate judgments of distance on a lawn, but inaccurate judgments in a lobby, and inconsistent and inaccurate judgments in a hallway.[30] Their reliability appears to be significantly affected by their location (outdoors or indoors).

These studies and others lead Olin and Doris to conclude, "[T]he cumulative evidence that cognitive functioning is contextually variable—make that *enormously* contextually variable—seems to us irrefutable."[31] And unreliability is just what we would expect given such variability—especially because the variability is often due to epistemically irrelevant factors. (We would expect epistemically irrelevant factors to bear a tenuous relationship to the truth.) Further, we have direct evidence that in the lawn/lobby/hallway studies, variation resulted in diminished performance. Whereas subjects perform reliably

---

[24] See Zagzebski (1996), chapter 4.

[25] See Olin and Doris (2014, pp. 670–672), for a summary of these studies.

[26] Reber and Schwarz (1999); cf. Alter et al. (2007); Song and Schwarz (2008).

[27] Smith and Schaffer (1995).

[28] Alter and Oppenheimer (2006), study 1; cf. McGlone and Tofighbakhsh (2000); Shah and Oppenheimer (2007, 2008).

[29] Proffitt et al. (2003); Proffitt (2006).

[30] Lappin et al. (2006); cf. Witt et al. (2007).

[31] Olin and Doris (2014, p. 671); emphasis in original. Thus, they say, "[A] wealth of evidence indicates that human cognition is highly contingent on contextual variation, making the cognitive capacities of normal people quite *un*reliable" (Olin and Doris 2014, 670; emphasis in original).

outdoors, merely moving them indoors appears to make them less reliable. Additionally, and by way of indirect evidence, there's no reason to think that such factors as font color or ease of pronunciation are the sorts of items that track the truth of the relevant judgments. These considerations, Olin and Doris think, cast doubt on the claim that human cognitive faculties are reliable.[32]

III. The Charges

Let's now consider how Strong Situationism-$_{CHAR}$ and Strong Situationism-$_{REL}$, respectively, figure into arguments against responsibilist and reliabilist VE.

Alfano argues that strong conservative virtue responsibilism—or at least the version defended by Linda Zagzebski—is vulnerable to skepticism. His argument is stated as an inconsistent triad:

> (Non-skepticism): Most people know quite a bit.
> (Strong Conservative Responsibilism): Knowledge is true belief acquired and retained through responsibilist intellectual virtue.
> (Strong Situationism-$_{CHAR}$): Most people's conative[33] intellectual traits are not virtues because they are highly sensitive to seemingly trivial and epistemically irrelevant situational influences.[34]

Alfano argues that Strong Situationism-$_{CHAR}$, conjoined with Zagzebski's account of knowledge, implies the denial of non-skepticism. Given Strong Situationism-$_{CHAR}$, we must either go skeptical or give up the strong conservative responsibilist account of knowledge. But because non-skepticism is a "Moorean platitude" that most philosophers want to embrace, Alfano recommends that Zagzebski's account of knowledge (and similar accounts) be jettisoned.

Olin and Doris appeal to Strong Situationism-$_{REL}$ as a skeptical threat to any version of VE on which knowledge requires the possession of reliable belief-producing faculties. We may summarize their argument like this:

1. If human beings lack stably reliable cognitive faculties, then any version of VE on which knowledge requires stably reliable cognitive faculties is threatened by skepticism.
2. (Strong Situationism-$_{REL}$) Human beings lack stably reliable cognitive faculties.

---

[32] I regret that space has not permitted discussion of studies related to the reliability of human inference. For discussion see Alfano (2013, chap. 6); and Fairweather and Montemayor (2014).

[33] *Conative* intellectual traits are traits of the mind that involve one's will or motivations.

[34] Alfano (2012, p. 234). What I call Strong Situationism-$_{CHAR}$ is just what Alfano calls "epistemic situationism." And what I call "strong conservative VE" is what Alfano calls "classical responsibilism."

3. Thus, any version of VE on which knowledge requires stably reliable cognitive faculties is threatened by skepticism.

Virtue reliabilists (and some strong conservative virtue responsibilists[35]) are committed to (1) by their analyses of knowledge. On such analyses, knowledge requires that an agent arrive at a true belief by way of a reliable faculty. In support of (2), Olin and Doris supply a staggering array of studies exhibiting the unreliability of certain belief-forming process-types across a range of circumstances (see section II). The conjunction of (1) and (2) entails (3). But the cost of skepticism is too high: most philosophers will abandon theses that are shown to imply skepticism.

The skeptical arguments against strong conservative responsibilism and reliabilist VE are importantly related to a second charge: *empirical inadequacy*. To put the point briefly, critics suggest that if these views are true, then *prima facie*, we would expect responsibilist and reliabilist intellectual virtues to be widely distributed over the population. But according to Alfano and Olin and Doris, these expectations are unmet. The empirical data violate what the theories predict. Thus, single features of the views—their empirical commitments—give rise to the twin problems of skepticism and empirical inadequacy.

As Alfano and Olin and Doris note, there are ways for virtue epistemologists to wriggle out of problematic empirical commitments. For instance:

- Strong conservative responsibilists might adjust their theories so that intellectual virtues are construed as *local* traits, such as *creativity-when-in-a-good-mood* or *courage-when-faced-with-a-non-unanimous majority*.
- Virtue reliabilists might adjust their theories so that the faculty-virtues required for knowledge are not broad-scope virtues like good vision or reliable memory. Rather, reliabilist virtues can instead be cast as narrow-scope virtues, such as the capacity to recognize words in the upper-right quadrant of one's visual field, or the capacity to gauge the distance of nearby objects while standing outdoors.

Arguably, such moves can help virtue epistemologists avoid skepticism and empirical inadequacy. At risk, however, is their theories' normative appeal. The worry is that local virtues such as *creativity when in a good mood* aren't epistemically praiseworthy in the same way as such global traits as creativity, intellectual courage, open-mindedness, and the like—because the latter traits are stable enough to manifest across a

[35] E.g., Zagzebski (1996), who analyzes intellectual virtue partly in terms of a reliability component, and goes on to analyze knowledge and epistemic justification in terms of intellectual virtue.

wide range of situations.[36] Similarly, if reliabilist virtues are described as narrowly as *reliable-distance-judgment-for-nearby-objects-while-outdoors*, then they seem undeserving of their name—for they aren't cognitive *excellences*. While these "narrowing" strategies may enable virtue epistemologists to avoid skepticism and empirical inadequacy, they seem to do so at the price of normative paucity. That is, they reduce the extent to which the intellectual virtues are *excellences* that can provide a normative ideal and serve a regulative function. While one might sensibly aspire to (say) intellectual creativity or courage, no one ever aspired to *creativity when in a good mood or courage under favorable conditions* or *upper right quadrant recognition reliability*.

IV. Replies to Situationist Objections

If VE is to remain tenable in the face of situationist critiques, it must face up to the empirical data while avoiding both skepticism and normative paucity. How might that be done? Because the varieties of VE differ in their empirical commitments, we'll treat them separately, with special attention to these commitments.

A. STRONG CONSERVATIVE RESPONSIBILISM

Let's start by considering how the strong conservative responsibilist can parry the situationist attack. As noted above, Zagzebski is the most prominent exponent of such responsibilism. On her account, knowledge is a state of true belief arising from acts of intellectual virtue. The situationist critique appeals to psychological research that seems to show that most humans lack the relevant virtues, and thus—given Zagzebski's analysis—knowledge. So, Zagzebski's view seems to lead to skepticism.

The responsibilist's first attempt to deflect the skeptical charge may note that Zagzebski explicitly rejects the thesis that *acts* of intellectual virtue require *agents* who possess responsibilist virtues.[37] Zagzebski's account of knowledge requires acts of intellectual virtue and virtuous motivation, but it doesn't require that knowers *have* the virtues. It is thus compatible with the

[36] Alfano (2012) emphasizes this point in several places.

[37] Thus, she says, "[O]n my definition of an act of virtue, it is not necessary that the agent actually possess the virtue. But she must be virtuously motivated, she must act the way a virtuous person would characteristically act in the same circumstances, and she must be successful because of these features of her act. What she may lack is the entrenched habit that allows her to be generally reliable in bringing about the virtuous end. This definition permits those persons who do not yet fully possess a virtue but are virtuous-in-training to perform acts of the virtue in question" (Zagzebski 1996, 279.)

conjunction of Non-skepticism and Strong Situationism-$_{CHAR.}$[38] After the first round of the dialectic, the skeptical threat seems to have dissipated.

But situationist critics can strike back.[39] Zagzebski's requirement of *acts of intellectual virtue* may not be as empirically thin as it appears—at least on the supposition that there are enough acts of intellectual virtue for her view to avoid skepticism. Here's why. In order for Zagzebski's analysis to avoid skepticism, there must be enough acts of intellectual virtue for it to be true that *most of us know quite a bit.* But now the critic can legitimately ask, "If we're not intellectually virtuous, what is it that enables most of us to perform a lot of intellectually virtuous acts?"[40] As Alfano notes, the most obvious answers aren't likely favorable news for the strong conservative responsibilist. Here are some possible answers, along with their attendant problems:

- *Suggestion 1:* Perhaps many acts of intellectual virtue are attributable to epistemic luck. *Problem*: This seems too convenient. On this suggestion, we'd have to think that most of us are *very lucky*. Further, certain varieties of epistemic luck are commonly thought to destroy knowledge rather than enable it.[41]
- *Suggestion 2:* Perhaps many acts of intellectual virtue are attributable to the fact that epistemic agents are *typically* in situations and moods that foster the exercise of "local" virtues, such as *creativity while in a good mood*. *Problem:* it is absurd to claim that epistemic agents are typically in virtue-friendly moods and situations. It is empirically obvious that cognitive agents traverse a wide range of moods and situations.
- *Suggestion 3:* Perhaps, despite their lack of responsibilist virtue, most people are motivated in the ways the virtuous person would be motivated, and act in ways the virtuous person would act. *Problem:* This seems like a disguised way to say that most people have responsibilist virtues. But in that case, the suggestion is empirically committed in a way that seems at odds with the empirical data, and strong conservative responsibilism turns out not to square with Strong Situationism-$_{CHAR}$, after all.

---

[38] See King (2014a); Miller (2014b); and Turri (forthcoming) on this point.

[39] One way of doing so is to note that acts of intellectual virtue, on Zagzebski's account, require *motivation* for epistemic goods. But perhaps empirical research will provide evidence that most of us lack the needed motivation. If that were to occur, Zagzebski's responsibilism may court skepticism even if it doesn't require virtue possession for knowledge. See King (2014a) for further discussion of this point.

[40] Alfano (forthcoming) develops this point in detail, and considers several of the suggestions discussed in this section.

[41] In order to make significant trouble for Suggestion 1, those pressing this worry would need to show that the variety of epistemic luck involved in such acts of intellectual virtue is epistemically malignant—for some varieties of epistemic luck are compatible with knowledge. See Pritchard (2005) for a seminal study of epistemic luck.

The strong conservative responsibilist who follows Zagzebski must affirm that most of us perform many acts of intellectual virtue, but must do so in a way that allows her to avoid skepticism given her other commitments. As Alfano suggests, this may not be as easy as it first appears.

What's a responsibilist to say? Here I'll sketch two possible replies that seem to me neither clearly compelling nor clearly wrong. The first starts with an observation regarding Suggestion 2: it needn't be true that most of us are *typically* in moods and situations that foster acts of intellectual virtue for it to be true that many of us perform *quite a few* acts of intellectual virtue. So long as we're in such moods and situations fairly often, it may turn out that most of us perform quite a few acts of intellectual virtue, and thus, that most of us know quite a bit (by responsibilist lights), even if we're not actually virtuous. This is compatible with the idea that most of us experience a wide range of moods and situations, many of which foil our ability to perform acts of intellectual virtue. To illustrate: suppose I'm a cognitive agent who lacks responsibilist virtues and that I'm hardwired so that my moods typically run contrary to those of an intellectually virtuous agent. Worse still, suppose my circumstances are more often than not inimical to my performing acts of intellectual virtue (I'm often inundated with electronic distractions, social pressure, taxing puzzles, and so on). I might nevertheless be able to perform enough acts of intellectual virtue for it to be true that I know a lot. Perhaps several times a week—when I'm compiling my grocery list or walking the college quad or reading the newspaper—my circumstances are friendly enough to foster the right motivations and allow me to perform intellectually virtuous acts. Over time, these rare occasions help me build up a large stock of responsibilist-certified knowledge.

If this is right, it provides the strong conservative responsibilist with an explanation for how we perform many acts of intellectual virtue while acknowledging that responsibilist intellectual virtues are rare. But situationists may retort that this strategy comes at a price: acts of intellectual virtue performed only in favorable conditions may seem unimpressive. They seem at least partly, and perhaps significantly, attributable to the situations themselves and not to cognitively praiseworthy features of the agents who perform them—at least on the hypothesis that such agents lack intellectual virtue.[42] It strains language to label the relevant behaviors "acts of intellectual virtue." So this way of advocating Suggestion 2 may avoid skepticism, but it may do so at the cost of reducing the normative significance of intellectually virtuous acts.

In light of potential problems with this strategy, let's consider another:

- *Suggestion 4:* Virtue responsibilists insist that intellectual virtues come in *degrees*. Perhaps very few agents possess global virtues to the fullest degree. Indeed, perhaps few of us possess the intellectual

[42] Alfano (2012) and Olin and Doris (2014) press this point.

> virtues to such a degree that they make a strong showing in empirical studies. But perhaps we nevertheless possess the virtues to a significant, even praiseworthy degree. (We're genuinely intellectually courageous in many contexts, but not courageous enough to pass Asch's tests; we're creative in many contexts, but not creative enough to solve the difficult Duncker candle task; and so on.) Thus, perhaps we have intellectual virtues to the degree needed to perform many acts of intellectual virtue, but not to the degree needed to perform well in empirical studies that set a "high bar" for intellectually virtuous behavior.

This reply, though not decisive, has two things going for it. First, since the notion of degrees of intellectual virtue is already a feature of virtue responsibilism independent of the situationist objection, the appeal is not ad hoc. Second, the reply is at least consistent with the empirical data. The studies by Asch, Isen, and Sherif, for instance, seem to test for a high degree of virtue. These studies thus seem ill-suited to show that people lack lower or middling degrees of virtue—even praiseworthy degrees.

Of course, one might worry that the appeal to degrees of virtue purchases anti-skepticism and empirical adequacy at the cost of normative paucity. Whether it's a low or middling degree of global virtue that's allegedly possessed, or just virtue-in-the-situation, one might think it is not praiseworthy—again due to the effects of candy or comedy or confederates who enable one to act (say) courageously.

Whatever its initial appeal, we should tread cautiously in endorsing this line of thought. Consider the inference from claims like

So-and-so was enabled to act virtuously in part due to her mood

to claims like

So-and-so acted virtuously *because of* her mood,

and especially from the latter to

So-and-so isn't virtuous to a praiseworthy degree.

No standard inference rule, inductive or deductive, licenses these inferences. To be sure, it is clear that subjects who only solve the Duncker candle task in the positive affect condition aren't as praiseworthy as those who pass in the control condition. But still: if someone is a pack of Smarties away from solving such a difficult problem, it seems harsh to say she lacks a praiseworthy degree of global virtue. Consider an analogy. Suppose your friend can bench press 300 pounds, but only if he drinks a Red Bull energy drink. When he raises the weight cleanly, are you tempted to say that he succeeded in lifting the weight *because* of the Red Bull? Maybe—if "because of" means

"in part because of." But are you tempted to say further that your friend isn't *strong*, but perhaps only *strong-when-having-ingested-Red-Bull*? Probably not—especially if your friend can bench 290 or even 250 without ingesting the beverage. The fact that the energy drink helped your friend lift 300 pounds would not automatically license the claim that he's not strong. Much would depend on other factors. Crucially, much would depend on how your friend fared with respect to other tests of strength. But if that's the way things go with respect to athletic feats, why should they be different with respect to tests for intellectual virtue?[43]

B. VIRTUE RELIABILISM

Recall the Olin-Doris argument discussed earlier:

1. If human beings lack stably reliable faculties, then any version of VE on which knowledge requires reliable cognitive faculties is threatened by skepticism.
2. (Strong Situationism-$_{REL}$) Human beings lack stably reliable faculties.
3. Thus, any version of VE on which knowledge requires stably reliable cognitive faculties is threatened by skepticism.

There is much to explore here, as the fast-growing literature on reliabilist VE and situationism attests.[44] I offer a modest contribution to the discussion by further distinguishing between varieties of Strong Situationism-$_{REL}$. Depending on the version, the view will be more or less plausible, its support from empirical studies will be more or less extensive, and it will be more or less poised to pose a skeptical threat to reliabilist VE.[45]

Here, as elsewhere, quantifiers are salutary. Strong Situationism-$_{REL}$ admits of at least three distinct readings. One possible reading is:

(Super-Duper-Strong Situationism-$_{REL}$): Human cognitive faculties are never reliable.

No proponent of the situationist critique would advocate this claim. It is obviously false (insert your favorite counterexample), self-defeating (if human cognitive faculties are never reliable, then neither are the faculties that produced

[43] My discussion of strategies by which strong conservative responsibilists can deflect the situationist charge should not be taken to imply that I endorse such views *qua* accounts of knowledge or justified belief. To my mind, the views face pressing objections on other fronts. See Alston (2000); Greco (2000b); and Baehr (2008, 2011).

[44] See Olin and Doris (2014); Alfano (2013, chap. 6); Alfano (forthcoming); Fairweather and Montemayor (2014); and Pritchard (2014).

[45] Note that reliabilist virtue epistemology is not the only view potentially under pressure from empirical evidence of our unreliability. That empirical evidence is relevant to *any* account of knowledge (e.g., *simple reliabilism*) that requires such reliability.

the claim itself), and without empirical support. To top it all off, the view has a dreadful name.

Moving on, consider:

(Pretty Strong Situationism-$_{REL}$): Human cognitive faculties are significantly less reliable than folk psychology would estimate.

This is the version of Strong Situationism$_{-REL}$ that is best supported by our empirical evidence. But what follows from the view? It doesn't follow that we're so unreliable that we can't know a lot by virtue reliabilist lights. We could be significantly less reliable than folk psychology would estimate, yet still be reliable in many circumstances (and thus have a lot of knowledge by virtue reliabilist lights). In short, this version of situationism is ill-suited to ground a skeptical threat to virtue reliabilism—it's just not strong enough to do skeptical lifting.

A final reading of Strong Situationism-$_{REL}$ lies between the previous two:

(Super-Strong Situationism-$_{REL}$): Human cognitive faculties are only very rarely reliable.

This claim, if true, would spell skeptical trouble for virtue reliabilism. For if knowledge requires that we possess and exercise reliable faculties and we possess or exercise these only very rarely, then it follows that we have very little knowledge. Super-Strong Situationism-$_{REL}$ is strong enough to push us from virtue reliabilism to skepticism. However, this view runs into trouble over its empirical support. It is also vulnerable to a problem of self-defeat. Let's take these problems in turn.

Suppose we take at face value the studies that show that we're unreliable with respect to some propositions under certain conditions. As critics of reliabilist VE themselves note, it doesn't follow that we're almost always unreliable. We may be reliable in many more "friendly" epistemic contexts. However, here the virtue reliabilist encounters a dilemma. Olin and Doris explain:

> The virtue epistemologist faces a decision: specify the domains [in which our faculties are reliable] *broadly* or specify domains *narrowly*. A broad specification counts a large and diverse range of circumstances as relevant to the functioning of a given epistemic virtue, while a narrow specification is restricted to a smaller, more uniform range. The broad horn appears to enjoy greater normative appeal. . . . However, there's trouble on the empirical side, for the attribution of broadly specified virtues is compromised by evidence of cognitive unreliability . . . but . . . If the empirical threat is ameliorated by indexing to very narrow domains, the resultant conception of virtue seems normatively slight (678–579).

Some virtue reliabilists seem poised to grasp the narrow horn of the dilemma (the horn that specifies reliable domains narrowly).[46] I'll neither endorse nor reject this move here. Instead, I'll explore a strategy that involves the broad horn—the horn that allegedly forces us back to empirical research showing that our faculties are unreliable.

Consider *vision*, broadly construed. And now consider the hallway studies. Suppose they show that most of us are unreliable at judging distances in hallways. It doesn't follow that our visual processing systems are unreliable—*even in hallways*. To see this, suppose I'm terrible at discerning whether an object in a hallway is (say) ten feet away or twelve. I might nevertheless be perfectly reliable with respect to the following claims, under the same conditions:

- I'm in a hallway.
- The hallway comprises walls and a floor.
- I'm not outdoors or in a basketball arena or a bar.
- There is (is not) a door to my right.
- There is an object before me.
- The object is (is not) red.
- The object is (is not) a beer bottle.
- The object is more than two feet away, and fewer than fifty.

And so on. Presumably, even though I'm unreliable with respect to the distance judgment that the lab coordinators assign as the target proposition, I'm reliable with respect to the propositions above. My visual inputs can be counted on to yield true output-beliefs. And if so, on virtue reliabilism, my visual faculty (even broadly construed) still enables me to know a lot, notwithstanding the fact the empirical research reveals an important respect in which I'm unreliable.

Crucially, this line of thought can be repeated with respect to other faculties (e.g., reasoning and memory, along with other sensory faculties, such as hearing and touch). Empirical research identifying the unreliability of those faculties vis-à-vis certain propositions under certain circumstances will not show that those faculties fail to yield knowledge about *other* propositions in the same conditions. This significantly undermines the support that Super-Strong Situationism-$_{\text{REL}}$ receives from the psychological research. That is to say, human cognitive faculties appear remarkably reliable over wide ranges of propositions even in those contexts where studies show us to be cognitively vulnerable in some respects. The unreliability claim doesn't generalize far enough to generate a skeptical threat.[47]

---

[46] See, e.g., Turri (forthcoming).

[47] Thanks to Jason Baehr and Christian Miller for helpful discussion here.

Let's turn to a further underminer for Super-Strong Situationism-$_{REL}$. If this claim is true, then human cognitive faculties are only very rarely reliable. But anyone who thinks that thereby has a *prima facie* reason to suspect that any belief she holds was unreliably formed. She therefore has reason to doubt Super-Strong Situationism-$_{REL}$ itself. Further, she has reason to doubt that the empirical research supporting the view was reliably conducted. And she has reason to doubt the reliability of her inference from the empirical data to Super-Strong Situationism-$_{REL}$. Perhaps such *prima facie* defeaters can be overcome, but at this stage in the game, the burden is on the situationist to show how this might be done. Those who wish to defend Super-Strong Situationism-$_{REL}$ will need to address this problem.

To sum up: it appears that Pretty Strong Situationism-$_{REL}$ enjoys significant empirical support, but is ill-suited for use in an argument that sticks virtue responsibilism with the skeptic's bill. And the stronger versions of Strong Situationism-$_{REL}$ face problems of self-defeat and insufficient empirical support. Thus, while virtue reliabilists should remain attuned to insights from empirical psychology, they need not feel threatened by the empirically based skeptical attacks that have been launched so far.

C. OTHER VARIETIES OF RESPONSIBILIST VE

To this point, we've focused on strong conservative responsibilist VE and reliabilist VE. Let's turn to the remaining species of responsibilism:

- *Weak conservative responsibilism*: responsibilist virtue concepts can supplement the treatment of traditional epistemological issues, such as the criteria, extent, and sources of our knowledge and justified belief. However, responsibilist virtues aren't essential to solving traditional epistemological problems.
- *Weak autonomous responsibilism*: responsibilist virtue concepts can prove helpful in addressing problems and issues that, while outside the scope of traditional epistemology, are nevertheless epistemological. VE is largely *independent* of (but compatible with) traditional epistemology.
- *Strong autonomous responsibilism*: responsibilist virtue concepts cannot be employed to address traditional epistemological problems, even in part. However, responsibilist virtues point to cognitive ideals and research programs that should *replace* those of traditional epistemology.

Note first that none of the above views implies that responsibilist virtues are central to the analysis of knowledge. Thus, as far as these views are concerned, it could turn out that few of us possess responsibilist virtues, while most of us nevertheless possess wide swaths of knowledge. In short, these

versions of responsibilism don't face a *skeptical* threat from the situationist critique.

How about the empirical inadequacy charge? Start by noting that our three versions of responsibilism do not, by themselves, carry empirical commitments that are at odds with any sort of Strong Situationism-$_{CHAR}$, or with the empirical research itself. These versions of responsibilism, just as such, take no stand about the distribution of responsibilist virtues. This point counsels caution about the idea that all versions of responsibilist VE are ipso facto subject to the situationist critique.

Are the remaining varieties of responsibilism off the hook? Not automatically. For even if the *views* as such don't conflict with the empirical research, the *projects* undertaken by proponents of responsibilist VE may still court such conflict. Further, proponents of the above views never hold them in the abstract. Instead, they hold them because the views fit with some particular virtue-oriented project. One doesn't typically start out as a certain kind of responsibilist, and then seek out particular responsibilist projects. In most cases, the order of explanation is the opposite. In order to see how our three varieties of responsibilism fare with respect to the situationist critique, we'll need to examine the empirical commitments involved in the relevant projects.

For this task, a representative sampling will have to do. We'll briefly consider one project characteristic of weak conservative responsibilism, weak autonomous responsibilism, and strong conservative responsibilism. In each case, we'll see that the projects demand only modest empirical commitments, so that they are compatible both with the extant psychological research and with some varieties of Strong Situationism-$_{CHAR}$. This is important. While most of the attention in virtue epistemology has gone to strong conservative responsibilism and virtue reliabilism, perhaps the majority of possible VE projects lie outside these categories. Thus, even if strong conservative and reliabilist VE are in trouble, it would not follow that VE itself should be abandoned. Rather, the bulk of the remaining projects could continue unhindered.[48]

1. Baehr's Weak Conservative Responsibilism

Jason Baehr (2008, 2011) has argued that something like the exercise of responsibilist virtues is needed to supplement reliabilist accounts of "high-grade" knowledge and evidentialist accounts of epistemic justification. Without appeal to responsibilist virtues, he thinks, such accounts fail to state *sufficient* conditions for their target concepts.

---

[48] Olin and Doris acknowledge this point (2014, note 7), but do not elaborate. By virtue of his aim to provide a situationist challenge to "all versions" of responsibilism, Alfano seems at least initially committed to denying that there are species of responsibilism that can steer clear of the empirical research (2012, p. 225).

Baehr eschews the project of giving intellectual character virtues a central role in perfectly general analyses of ordinary knowledge and non-reflective, non-evidential varieties of epistemic justification. He grants that much "low-grade" knowledge can be achieved in just the way reliabilists suggest. Likewise, he grants that there are species of epistemic justification that do not require the correct (and virtuous) assessment of evidence. In this respect, the claims he advocates are considerably weaker than those of (e.g.) Zagzebski, who provides perfectly general analyses of ordinary epistemic properties. Baehr's aim is simply to show that analyses of certain epistemic statuses require supplementation via intellectual character virtues.

Baehr's supplementation of reliabilism and evidentialism might nevertheless incur empirical costs. If "high-grade" knowledge and certain varieties of epistemic justification require the exercise of intellectual virtue, then situationist data indicating that most of us lack intellectual virtue will imply that few of us achieve these epistemic statuses. By way of brief reply, note first that it's not a Moorean platitude that many of us possess a lot of "high-grade" knowledge and reflective, virtuously achieved, evidence-based epistemic justification. So wherever Baehr's analyses lead us, they appear not to lead us into *skepticism*—at least not skepticism about many varieties of knowledge and justification. Further, Baehr's accounts could be amended so that, like Zagzebski's, they require *acts characteristic of intellectual virtue*, rather than *exercises* of intellectual virtue, for high-grade varieties of knowledge and justification.

These admittedly brief comments suffice to show that Baehr's weak conservative responsibilism carries lean empirical commitments. The commitments it does carry *may*, conjoined with the findings of empirical psychology, imply that certain exalted epistemic states are rarely realized. But we might have expected as much anyway. That's about as bad as the news can get. Baehr's projects don't court skepticism about ordinary knowledge or justification, and it is by no means clear that they court empirical inadequacy. It seems, therefore, that at least some lines of weak conservative responsibilist inquiry can coexist peacefully with the extant empirical research.

2. Roberts and Wood: Weak Autonomous Responsibilism

Robert Roberts and Jay Wood (2007) undertake a responsibilist project intended to provide a conceptual map of the intellectual virtues. The authors begin by locating their project in relation to traditional epistemology. They indicate no desire to provide a rigorous definition of knowledge. Indeed, they chide epistemologists caught up in post-Gettier (1963) attempts to patch, puncture, and replace the traditional justified true belief account of knowledge.

Though they don't reject traditional epistemology altogether, Roberts and Wood defend an independent kind of epistemology intended to offer both

lofty cognitive ideals and valuable cognitive guidance. Their favored brand of epistemology features detailed characterizations of such intellectual virtues as courage, caution, firmness, charity, generosity, and open-mindedness. The relevant characterizations typically include inspiring biographies of exemplars of the relevant virtue. For instance, Jane Goodall's persistence in studying an unpredictable and dangerous species of chimpanzees makes her an exemplar of intellectual courage.[49] Roberts and Wood go on to provide analyses of the virtues, and employ their exemplars to glean insights for the analyses. (For example, Goodall's resistance to *fears* and *threats* in her exercises of courage reveals that courage involves managing fears and properly responding to threats.) What emerges from their project, Roberts and Wood hope, is an understanding of responsibilist virtues that clarifies the nature of these virtues and the relations between them, along with inspiration to seek the virtues and their attendant epistemic goods.

Central to their project is the goal of conceptual clarity regarding responsibilist virtues. Such a pursuit does not require the empirical claim that responsibilist virtues are widely distributed. Indeed, among the first conceptual points Roberts and Wood make about virtues is that they are *excellences:* "We propose that in general a human virtue is an acquired base of excellent functioning in some generically human sphere of activity that is challenging and important."[50] Given this, one might think that if their responsibilism predicts anything, it predicts that responsibilist virtues will be rare. Finding the virtues to be so might actually *confirm* this sort of responsibilism.

But perhaps Roberts and Wood don't have a project wholly devoid of empirical commitments. One might think that their project requires at least a few real-life exemplars in order to support their characterizations of individual virtues. That would be a modest empirical commitment, but a commitment nonetheless. But, on reflection, their project may not even require this much. If we were to learn, for instance, that Goodall is a fiction, thinking about her story would still teach us about courage. Indeed, one might think, if exemplars did not exist, we'd have to invent them. What's most important is the *role* exemplars play. According to Roberts and Wood, the exemplars of intellectual virtues, along with the corresponding analyses, function as cognitive ideals that can perform an important normative function: namely, they give us actions to emulate and characters to strive for. And we don't need *actual* exemplars for that. So for this portion of their project, virtue responsibilists require only very weak empirical commitments to the effect that someone could have had such-and-such a character. Clearly, no psychological study threatens such claims.[51]

---

[49] See Roberts and Wood (2007), 145–8 and 223–4.

[50] Roberts and Wood (2007), 59.

[51] Thanks to Nathan Ballantyne for helpful discussion here.

One further respect in which Roberts and Wood appear empirically committed concerns the possibility of our *acquiring* intellectual virtues, or at least emulating virtuous exemplars in such a way that we perform acts of intellectual virtue. And perhaps such a commitment renders their view vulnerable to situationist attack. This worry is similar to one raised below in connection with Kvanvig's strong autonomous responsibilism, so we'll address it there. [52]

3. Kvanvig: Strong Autonomous Responsibilism

Jonathan Kvanvig (1992) argues that epistemic properties such as knowledge and justification cannot be analyzed in terms of intellectual virtue. In his view, the problems and projects of traditional epistemology leave little room for intellectual virtue concepts. But while some epistemologists might see this as a reason to jettison the concept of an intellectual virtue, Kvanvig takes it as his mandate to reject much of traditional epistemology and assert a virtue-centric alternative.

Let's set aside Kvanvig's arguments that much of traditional epistemology should be rejected and focus instead on his positive proposal: that one goal of epistemology is to provide advice that enables us to achieve true belief, avoid error, and develop the intellectual virtues.[53] This side of the project can be separated from the side that rejects traditional epistemology. (Indeed, *any* virtue epistemologist can accept the positive side of Kvanvig's project, provided it is suitably recast as a supplement to traditional epistemology.)

On Kvanvig's view, "[w]hat we really want from an epistemologist is an account of the cognitive life of the mind that addresses our cognitive experience and helps us understand how to maximize our potential for finding truth and avoiding error" (vii). For Kvanvig, the move toward such an account requires a paradigm shift from an individualist, Cartesian picture of epistemology to a socially oriented epistemology that both accounts for the communal aspects of knowledge and features an emphasis on the intellectual virtues. Central to Kvanvig's project, then, is the idea that epistemology should be reoriented so that it serves a regulative function: helping *us* together achieve a greater proportion of true beliefs. Kvanvig emphasizes that we must be raised in the right kind of intellectual setting in order to possess a range of intellectual goods, including knowledge, intellectual virtue, and appropriate attitudes toward skepticism. Part of the relevant kind of socialization involves the emulation of exemplars of intellectual virtue.

---

[52] Though the evaluation of weak autonomous responsibilism in this section has focused primarily on the work of Roberts and Wood, there are several other proponents of the view. See, e.g., Code (1987) and Hookway (2003). Of course, the relevant projects must be individually assessed with respect to the situationist challenge. Not all versions of weak autonomous responsibilism carry the same empirical commitments. The point of the present section has just been to show that this version of responsibilism *need not* require ambitious empirical claims.

[53] See Baehr (2008) for critical discussion of Kvanvig's arguments against traditional epistemology.

Like the weak autonomous responsibilism developed by Roberts and Wood, Kvanvig's brand of responsibilism carries lean empirical commitments. First, at least as Kvanvig explicates it, it requires there to be exemplars of intellectual virtue for the rest of us to emulate. Second, it requires that those of us who don't currently possess these virtues can come to improve our proportion of true beliefs to false ones—perhaps via exercises or actions characteristic of intellectual virtue. Thus, Kvanvig seems committed to the claim that we can progress toward intellectual virtue.

As we've seen, the first commitment is modest—exemplars of intellectual virtue needn't be common (or even actual) in order to play their regulative role. Further, it is plausible that there are at least some exemplars of responsibilist virtues to be found.[54] At any rate, such an empirical commitment is in keeping with even a "situationist-friendly" assessment of the empirical data. What about the second commitment? Here again, we find a claim that is compatible with the idea that very few of us possess the responsibilist virtues. Kvanvig's project requires that we *can* acquire responsibilist virtues, but needn't presuppose much about the actual distribution of these traits across the population. The psychological evidence indicates *at most* that very few of us currently possess responsibilist virtues. But the fact that we don't *currently* possess these virtues is weak evidence for the claim that such virtues are unattainable for us. (Compare: that very few preschoolers can read, write, and do basic arithmetic is weak evidence for the claim that they can't learn to do so proficiently. The prospect of acquired skills is a fundamental assumption of the educational process. Why then dismiss a similar prospect in the case of intellectual virtue inculcation?) At any rate, by virtue of its modest empirical commitments, Kvanvig's version of strong autonomous responsibilism squares with the available data, thereby avoiding the situationist critique.[55]

V. Toward Collaboration

Having considered a number of situationist challenges to various kinds of virtue epistemology, we've seen that, on the whole, VE is compatible with the findings of empirical psychology. To sum up:

- Strong conservative VE needn't involve commitment to there being many robustly virtuous cognitive agents. Its accounts of knowledge

[54] At least *prima facie*, Roberts and Wood (2007) and Baehr (2011) are replete with accounts of actual exemplars. See also King (2014b).

[55] A similar point applies to the claim that a central goal of education should be the inculcation of responsibilist virtues. For detailed discussion of this point, see Battaly (2014) and Baehr (forthcoming). See Miller (2014a), chapters 8 and 9, for discussion of related issues regarding the inculcation of moral virtue.

and justified belief require only that there be many *acts* of intellectual virtue.

- Strong conservative VE faces an explanatory task of accounting for there being many acts of intellectual virtue without there being many robustly virtuous agents. One way in which this task might be accomplished is by appeal to the independently motivated concept of *degrees of virtue*.
- Reliabilist VE appears poised to fend off the situationist attack. The empirical research does not support forms of situationism about our reliability that are strong enough to saddle the virtue reliabilist with skepticism.
- The remaining varieties of VE are modest in their empirical commitments, and do not require acts or exercises of intellectual virtue for knowledge or justified belief. Such species of VE thus steer clear of skeptical threats from situationism and empirical inadequacy.

If these conclusions are right, many VE projects can coexist peacefully with empirical psychology, and even with several versions of situationism.

But why settle for peaceful coexistence when there are significant opportunities for friendly collaboration? The latter are worth exploring. For example, virtue reliabilists should eagerly assimilate data that reveals contexts in which our cognitive faculties are unreliable with respect to certain propositions.[56] Such data can both improve our estimations about the extent of our knowledge and forewarn us of epistemically malign circumstances. Similarly, virtue responsibilists can and should mine the psychological literature for insights about the situations that make us more (or less) likely to act characteristic of intellectual virtue. Such insights, coupled with regulative ideals gleaned from epistemic exemplars and analyses of individual intellectual virtues, might offer significant cognitive guidance.

These suggestions are admittedly speculative. And given their speculative status, the suggestions do not call for an arranged marriage between virtue epistemology and empirical psychology. But given what might be gained via collaboration between these approaches, it seems reasonable to suggest a trial friendship.[57]

---

[56] For a model of how this might look, see Alvin Goldman's extensive interaction with empirical psychology in Goldman (1986). For recent work joining together themes from epistemology with psychological research see Kornblith (1999), Kelly (2008), and Ballantyne (2015).

[57] Thanks to Jason Baehr, Nathan Ballantyne, and Christian Miller for insightful comments and helpful discussion. This publication was made possible through the support of a grant from The Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. Finally, thanks to Whitworth University for providing research leave for this project.

References

Alfano, Mark. (2012). "Expanding the Situationist Challenge to Responsibilist Virtue Epistemology." *Philosophical Quarterly* 62(4): 223–49.

———. (2013). *Character as Moral Fiction*. Cambridge: Cambridge University Press.

———. (Forthcoming). "Epistemic Situationism: An Extended Prolepsis." In Alfano and Fairweather (eds.), *Epistemic Situationism*. New York: Oxford University Press.

Alfano, Mark, and Abrol Fairweather, eds. (Forthcoming). *Epistemic Situationism*. New York: Oxford University Press.

Alston, William P. (2000). "Virtue and Knowledge." *Philosophy and Phenomenological Research* 60: 185–189.

Alter, Adam L., and Daniel M. Oppenheimer. (2006). "Predicting Short-Term Stock Fluctuations by Using Processing Fluency." *Proceedings of the National Academy of Science*s 103: 9369–9372.

Alter, Adam L., Daniel M. Oppenheimer, Nicholas Epley, and Rebecca N. Eyre. (2007). "Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning. *Journal of Experimental Psychology* 136: 569–576.

Asch, Solomon. (1951). "Effects of Group Pressures upon the Modification and Distortion of Judgment." In Harold Guetzkow (ed.), *Groups, Leadership, and Men*, 177–190. Pittsburgh, PA: Carnegie Press.

———. (1952). *Social Psychology*. New York: Prentice-Hall.

———. (1955). "Opinions and Social Pressure." *Scientific American* 193 (5): 31–35.

———. (1956). "Studies of Independence and Conformity: A Minority of One against a Unanimous Majority," *Psychological Monographs* 70(9): 1–70.

Baehr, Jason. (2008). "Four Varieties of Character-Based Virtue Epistemology." *Southern Journal of Philosophy* 46(4): 469–502.

———. (2011) *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. New York: Oxford University Press.

———. (Forthcoming) Manuscript on situationism and educating for intellectual virtue. In Alfano and Fairweather (eds.), *Epistemic Situationism*. New York: Oxford University Press.

Ballantyne, Nathan. (2015). "Debunking Biased Thinkers (Including Ourselves)." *Journal of the American Philosophical Association*, 1(1): 141–162.

Battaly, Heather. (2014). "Acquiring Epistemic Virtue: Emotions, Situations, and Education." In Fairweather and Flanagan, *Naturalizing Epistemic Virtue*. Cambridge: Cambridge University Press, 175–196.

Code, Lorraine. (1987). *Epistemic Responsibility*. Hanover, NH: University Press of New England and Brown University Press.

Doris, John. (1998). "Persons, Situations, and Virtue Ethics." *Noûs* 32(4): 504–30.

———. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Fairweather, Abrol, and Carlos Montemayor. (2014). "Inferential Abilities and Common Epistemic Goods." In Abrol Fairweather (ed.), *Virtue Epistemology Naturalized*. Dordrecht: Springer, 123–141.

Fairweather, Abrol, and Owen Flanagan, eds. (2014). *Naturalizing Epistemic Virtue*. Cambridge: Cambridge University Press.

Gettier, Edmund. (1963). "Is Justified True Belief Knowledge?" *Analysis* 23(6): 121–123.
Goldman, Alvin. (1979). "What Is Justified Belief?" In George Pappas (ed.), *Justification and Knowledge*, 1–23. Dordrecht: D. Reidel.
———. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
Greco, John. (2000). *Putting Skeptics in Their Place*. New York: Cambridge University Press.
———. (2000b). "Two Kinds of Intellectual Virtue." *Philosophy and Phenomenological Research* 60(1): 179–184.
———. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. New York: Cambridge University Press.
Greco, John, and John Turri (2011). "Virtue Epistemology." *The Stanford Encyclopedia of Philosophy*. Stable URL: http://plato.stanford.edu/entries/epistemology-virtue/
Harman, Gilbert. (1999). "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99: 315–331.
Hookway, Christopher. (2003). "How to Be a Virtue Epistemologist." In Michael DePaul and Linda Zagzebski (eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology*, 183–202. New York: Oxford University Press.
Isen, Alice, Kimberly Daubman, and Gary Nowicki. (1987). "Positive Affect Facilitates Creative Problem Solving." *Journal of Personality and Social Psychology* 52: 1122–1131.
Kelly, Thomas. (2008). "Disagreement, Dogmatism, and Belief Polarization." *Journal of Philosophy* 105: 611–633.
King, Nathan. (2014a). "Responsibilist Virtue Epistemology: A Reply to the Situationist Challenge." *Philosophical Quarterly* 64(255): 243–253.
———. (2014b). "Erratum to: Perseverance as an Intellectual Virtue." *Synthese*. 191(15): 3779–3801.
Kornblith, Hilary. (1999). "Distrusting Reason." *Midwest Studies in Philosophy* 23: 181–196.
Kvanvig, Jonathan. (1992). *The Intellectual Virtues and the Life of the Mind: On the Place of the Virtues in Contemporary Epistemology*. Savage, MD: Rowman and Littlefield Press.
Lappin, Joseph S., Amy L. Shelton, and John J. Reiser. (2006). "Environmental Context Influences Visually Perceived Distance." *Perception and Psychophysics* 68(4): 571–581.
McGlone, Matthew S., and Jessica Tofighbakhsh. (2000). "Birds of a Feather Flock Conjointly(?): Rhyme as Reason in Aphorisms." *Psychological Science* 11: 424–428.
Miller, Christian. (2013). *Moral Character: An Empirical Theory*. New York: Oxford University Press.
———. (2014a). *Character and Moral Psychology*. New York: Oxford University Press.
———. (2014b). "Moral Virtues, Epistemic Virtues, and the Big Five." In Fairweather and Flanagan (eds.), *Naturalizing Epistemic Virtue*, 92–117. New York: Cambridge University Press.
Olin, Lauren, and John Doris. (2014). "Vicious Minds: Virtue Epistemology, Cognition, and Skepticism." *Philosophical Studies* 168: 665–692.
Pritchard, Duncan. (2005). *Epistemic Luck*. New York: Oxford University Press.
———. (2014) "Re-evaluating the Situationist Challenge to Virtue Epistemology." In Fairweather and Flanagan (eds.), *Naturalizing Epistemic Virtue*, 143–154. Cambridge: Cambridge University Press.
Proffitt, Dennis R. (2006). "Embodied Perception and the Economy of Action." *Perspectives on Psychological Science* 1(2): 110–122.

Proffitt, Dennis R., Jeanine Stefanucci, Tom Banton, and William Epstein (2003). "The Role of Effort in Perceiving Distance." *Psychological Science* 14: 106–112.

Reber, Rolf, and Norbert Schwarz. (1999). "Effects of Perceptual Fluency on Judgments of Truth." *Consciousness and Cognition* 8: 338–342.

Roberts, Robert C., and Jay Wood (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*. New York: Oxford University Press.

Shah, Anuj K., and Daniel M. Oppenheimer. (2007). "Easy Does It: The Role of Fluency in Cue Weighting." *Judgment and Decision Making* 2: 371–79.

———. (2008). "Heuristics Made Easy: An Effort-Reduction Framework." *Psychological Bulletin* 134: 207–22.

Sherif, Mufazer. (1937). "An Experimental Approach to the Study of Attitudes." *Sociometry* 1: 90–98.

Smith, Stephen M., and David R. Schaffer. (1995). "Speed of Speech and Persuasion: Evidence for Multiple Effects." *Personality and Social Psychology Bulletin* 21: 1051–60.

Song, Hyunjin, and Norbert Schwarz. (2008). "Fluency and the Detection of Misleading Questions: Low Processing Fluency Attenuates the Moses Illusion." *Social Cognition* 26(6): 791–99.

Sosa, Ernest. (1980). "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge." *Midwest Studies in Philosophy* 5: 3–25.

———. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Vol. 1. New York: Oxford University Press.

———. (2009). *Reflective Knowledge: Apt Belief and Reflective Knowledge*. Vol. 2. New York: Oxford University Press.

Turri, John. (Forthcoming). "Epistemic Situationism and Cognitive Ability." In Alfano and Fairweather (eds.), *Epistemic Situationism*. New York: Oxford University Press.

Witt, Jessica K., Jeanine K. Stefanucci, Cedar R. Reiner, and Dennis R. Proffit (2007). "Seeing Beyond the Target: Environmental Context Affects Distance Perception." *Perception* 36, 1752–1768.

Zagzebski, Linda. (1996). *Virtues of the Mind*. New York: Cambridge University Press.

14 }

From Virtue Epistemology to Abilism: Theoretical and Empirical Developments

John Turri

Introduction

The basic idea behind what is nowadays called "virtue epistemology" is ancient. The basic idea is that we are endowed with certain cognitive powers, such as perception, memory, and inference. When we accurately represent and accept information through the operation of these powers, we have knowledge. That's what knowledge is. On my reading, this basic idea has been endorsed by Aristotle, Hume, Reid, Peirce, Russell, Sellars, and others in the history of philosophy. On the contemporary scene, the basic idea has been elaborated and defended in various ways, most notably by Linda Zagzebski (1996, 2009); Ernest Sosa (1991, 2007, 2011); John Greco (1993, 2010); and myself (2010, 2011). If contemporary virtue epistemology had to be boiled down to a simple slogan, it could be this: knowledge is true belief manifesting intellectual virtue. Intellectual virtues are defined as (some subset of the) reliable cognitive abilities or traits (for overviews of the field, see Greco & Turri 2011, 2012; for alternative takes on virtue epistemology, see Kvanvig 1992; Roberts & Wood 2007; and Baehr 2011).

Virtue epistemologists have identified three main theoretical benefits of their view. First, it explains why knowledge is better than mere true belief, which has been a central philosophical issue ever since Plato's *Meno* (for an overview, see Pritchard & Turri 2011). We value achievement over mere success, at least when the outcome is good or neutral. For example, compare a canvas painting whose pleasing lines manifest the painter's artistic powers to a painting whose pleasing lines are due to a janitor accidentally spilling some cleaning supplies on it (example adapted from Sosa 2003). The former is better than the latter. Similarly, compare a stretch of cognition whose eventual true belief manifests the agent's cognitive powers to a true belief due to a lucky guess or quirky intervention. The former is better than the latter (Riggs 2002).

Second, virtue epistemology places knowledge attribution in a familiar pattern, which is thought to be important because understanding the logic and purpose of knowledge attributions is arguably the most prominent theme in contemporary epistemology (e.g., Austin 1956; Stroud 1984; DeRose 1996; Turri 2013b). Whether it's art, athletics, oratory, or inquiry, we're keen to assess how outcomes relate to the relevant abilities or powers. Knowledge attributions are thus similar to other credit attributions (Greco 2003).

Third, virtue epistemology solves a stubborn problem about the relationship between knowledge and luck, "the Gettier problem," which analytic philosophers have discussed since the early 1960s (Gettier 1963; Zagzebski 1996; Greco 2003; Sosa 2007; for an overview of the problem, see Turri 2012a). Gettier cases are conjured by following a recipe (Zagzebski 1994; compare Sosa 1991: 238). Start with a belief sufficiently well formed that it would ordinarily qualify as knowledge. Then add an element of bad luck that would ordinarily prevent the well-formed belief from being true. Finally, add a dose of good luck that "cancels out the bad," so that the belief ends up true anyhow. (For an example that fits this structure, see section 3.1). Philosophers claim that the irresistible intuition is that the agent in such a case lacks knowledge. Virtue theory offers an explanation for why knowledge is absent: the true belief doesn't manifest the agent's cognitive powers. This fits seamlessly into a more general pattern of "manifestation failure," in which the outcome does not manifest the relevant power or ability (Turri 2011b, 2012b).

Several challenges face virtue epistemology. For example, it purports to explain certain intuitive facts about knowledge, but recent work has challenged the robustness and significance of the relevant intuitions (Hazlett 2010, 2012; Myers-Schulz & Schwitzgebel 2013; Murray, Sytsma & Livengood, forthcoming; Turri, forthcoming c; Starmans & Friedman 2012). Moreover, some charge that it presupposes an empirically false cognitive psychology (Alfano 2011; Olin & Doris 2014). And some complain that a key component of the theory, "manifestation," is uninformative, obscure, or intuitively inadequate to the task (e.g., Murphy 1988; Levin 2004; Roberts & Wood 2007; Pritchard 2009; Church 2013; Jarvis 2013; Kallestrup & Pritchard, forthcoming).

This chapter describes several theoretical and empirical results relevant to assessing these challenges and identifying areas for future research. Virtue epistemology, in its most popular and recognizable contemporary forms, does not survive the ordeal. Instead, what emerges is a leaner theory of knowledge that is more empirically adequate, better captures the ordinary conception of knowledge, and is ripe for cross-fertilization with cognitive science. I call this view *abilism*.

Truth, Belief, and Reliability

Leading virtue epistemologists defend the view that knowledge is *true belief* manifesting *reliable* intellectual virtue or ability (Zagzebski 1996; Sosa 2007; Greco 2010). But recent work in philosophy and psychology challenges the three italicized aspects of this theory of knowledge: truth, belief, and reliability.

Truth

Allan Hazlett (2010, 2012) argues that the ordinary concept of knowledge does not require truth. In other words, the ordinary concept of knowledge is not *factive*. In support of this view, Hazlett appeals to patterns in ordinary judgments about knowledge, which seem to involve natural and familiar non-factive usage (e.g., "At that moment, I knew I was about to die—but then I was saved"; "Medieval peasants knew that the Earth was flat, but now we know it wasn't"). Hazlett argues that a charitable interpretation of this data involves a non-factive conception of knowledge. Based on this, Hazlett proposes that knowledge is (roughly) reliably produced belief and that epistemologists should not concern themselves with ordinary usage or the ordinary concept of knowledge, because the philosophically interesting notion of knowledge is factive.

Virtue epistemologists could accept Hazlett's argument at face value and satisfy themselves with investigating knowledge in some technical sense that does require truth. But I seriously doubt that this will seem attractive to them. Rather, my sense is that they are interested in theorizing about the epistemic status we ordinarily think about, talk about, and value—namely, knowledge. For my part, this is what I'm interested in.

There are several more direct responses available. First, one might show that a charitable interpretation of the data needn't involve positing literal and competent applications of a non-factive conception of knowledge. (A competent application of the concept involves no egregious performance errors, and a literal application is just what it sounds like—no metaphor, irony, sarcasm, or the like.) I have done this on theoretical grounds (Turri 2011c), and recent experimental studies have yielded corroborating results (e.g., Buckwalter, under review). Second, one might show that knowledge ordinarily understood doesn't resemble reliably produced belief and, moreover, that truth is an extremely important part of knowledge ordinarily understood. Recent work supports this claim (Turri & Friedman 2014; Turri, 2014). Third, one might argue that even if knowledge is non-factive, we can still achieve the benefits of standard virtue-theoretic definitions of knowledge. I have developed this approach, arguing that it is, at worst, a very small concession to allow knowledge of literally false but approximately true claims (Turri, forthcoming a, forthcoming b). (An example of a false but approximately true claim might be

to say that a woman who is 1.98 meters tall is 2 meters tall.) Indeed, I suspect that allowing for such approximation better approximates the actual features of knowledge. No argument has ever been given for supposing that knowledge requires strict truth, rather than approximate truth.

The first amendment I propose to the standard virtue-theoretic definition, then, is this: knowledge is *approximately* true belief manifesting intellectual virtue.

Belief

For decades contemporary epistemologists have treated it as an intuitive datum that knowledge entails belief (e.g., Chisholm 1989). But this assumption is not uncontroversial in the history of epistemology (e.g., Radford 1966, Plato's *Republic* 477). And recently several philosophers have made a strong case that knowledge does not entail belief. Multiple experimental studies have shown that people are often perfectly willing to ascribe knowledge in the absence of belief (Myers-Schulz & Schwitzgebel 2013; Murray, Sytsma & Livengood, forthcoming). In some cases, the *vast majority* of people who attribute knowledge decline to attribute belief. Unless people are deeply confused about knowledge and belief, then these results provide strong evidence that knowledge does not entail belief. And if knowledge does not entail belief, then standard virtue-theoretic definitions of knowledge are false.

How might virtue epistemologists respond? They could accept the findings at face value and satisfy themselves with theorizing about knowledge in some technical sense that does require belief. But, as mentioned earlier, I am unaware of any support for this type of response among virtue epistemologists. In any event, I am uninterested in such a response. Virtue epistemologists could argue that the findings should be discounted because they are untrustworthy, the result of noisy data or suspect methodology. But the results have been replicated multiple times by different research teams using diverse stimuli and experimental designs. Virtue epistemologists could argue that the findings should be discounted because they result from error and confusion. If error and confusion are responsible, it runs deep and wide. But no evidence has been produced that this is happening. In any event, I strongly prefer to avoid positing an error theory of disturbing proportion.

Subsequent work has made considerable progress in explaining the concept of belief, why knowledge does not require belief, and, more importantly, what knowledge does require (Buckwalter, Rose & Turri, forthcoming; see also Rose, Buckwalter & Turri, 2014; Turri, Rose, & Buckwalter, under review; Buckwalter & Turri, under review a, under review b). Belief ordinarily understood is connected with *feeling* or *emotional commitment* in a way that knowledge is not. As William James wrote, "In its inner nature belief . . . is a sort of

feeling more allied to the emotions than to anything else" (James 1889: 21; see also Hume 1748 on belief as a *feeling*). It turns out that knowledge is not essentially connected with feeling in this way, which is why it does not—indeed, cannot—require belief. But even though knowledge does not require belief in this sense, there is every indication that it does require a "thinner" assertive representational state. This thinner state corresponds with what might ordinarily be described as *thinking that something is true*. If you know something, then although you might not *believe* it's true, you nevertheless must *think* it's true. For convenience, this thinner state has been stipulatively dubbed *thin belief*.

The second amendment I propose to the standard virtue-theoretic definition, then, is this: knowledge is approximately true *thin* belief manifesting intellectual virtue.

Reliability

Decades of research in social psychology taught us counterintuitive but valuable lessons about the determinants of human behavior. Situational factors influence our behavior to an extent that common sense wouldn't predict and which is shocking upon reflection (e.g., Hartshorne & May 1928; Milgram 1974; Darley & Batson 1973). Although people's behavior is fairly consistent over time in very similar situations, it can be highly inconsistent across situations that differ in ways that we might ordinarily think are insignificant (Mischel & Peake 1982). Moreover, the predictive value of situational variables can exceed the predictive value of (what we take to be) a person's traits such as honesty or generosity. Indeed, the predictive value of traits can be startlingly low (Ross & Nisbett 2011: 95).

It's been two decades now since philosophers began seriously coming to grips with the social psychological findings (Flanagan 1991; Doris 1998, 2002; Harman 1999). Gilbert Harman and John Doris clarified the findings' dramatic importance for ethical theory, especially traditional forms of virtue ethics which presuppose that people have character traits underwriting long-term, stable, and robust dispositions to be motivated and to act in particular ways. Do people have such character traits, such as honesty and compassion? Harman and Doris boldly suggested that decades of psychological science are relevant to answering this question, and they concluded that the science warrants a negative verdict.

Philosophers have recently extended the situationist challenge from virtue ethics to virtue epistemology (Alfano 2011; Olin & Doris 2014). They claim that cognitive psychology provides evidence that human cognition is highly contextually variable and, in many cases, unreliable. Just consider the litany of biases and foibles enumerated in textbooks on judgment and decision-making—the

availability bias, the confirmation bias, the anchoring bias, the false consensus effect, base-rate neglect, the conjunction fallacy, and so on. If situationists are right about this, then standard virtue-theoretic definitions of knowledge have implausible skeptical consequences. For these definitions say that knowledge requires intellectual virtue, and intellectual virtues must be reliable. (For an alternative reading of the relevant relationship between virtue and reliability, see Miller 2014.) If we aren't reliable, then we don't have virtues; and if we don't have virtues, then we don't have knowledge. In short, just as traditional virtue ethics presupposes an empirically inadequate moral psychology, it looks like (non-skeptical) virtue epistemology presupposes an empirically inadequate cognitive psychology.

I am not convinced that cognitive psychology shows what these situationists say that it shows (Turri, forthcoming b). But suppose that situationists dig deeper and convince us that our best cognitive science shows that we get things wrong more often than not. Certainly this is possible, and, in fact, I wouldn't be too surprised if it turned out to be true. Would we conclude that a wide-ranging skepticism is true?

I wouldn't, for two reasons. First, there is no evidence that knowledge must be reliably produced. Philosophers have simply assumed that knowledge requires reliability, without offering any evidence in support of this assumption. Second, there is good evidence that unreliably produced knowledge is not only possible but actual (Turri, forthcoming c, 2012c). On the one hand, lots of our knowledge is due to explanatory reasoning, and explanatory reasoning seems to get it wrong at least as often as it gets it right. On the other hand, many achievements in life much more impressive than knowledge don't require reliable abilities, so it stands to reason that knowledge doesn't require reliable abilities either. Knowledge needn't proceed from intellectual virtues. Instead, it can also proceed from cognitive powers or abilities that are less reliable and more common than intellectual virtues.

Some philosophers have suggested that "commonsense" or "folk" epistemology is implicitly a form of proto-reliabilism (Dretske 1981: 92, 249n8; Sosa 1991: 100; Goldman 1993: 271). If true, this proto-reliabilist hypothesis would provide some motivation for its namesake in the professional literature. But an actual investigation of commonsense knowledge judgments revealed strong evidence against the proto-reliabilist hypothesis (Turri, under review). It turns out that folk epistemology fully embraces the possibility of unreliable knowledge.

I accept the following metaphysical thesis about powers in general:

> If a person possesses an ability/power to produce an outcome (of a certain type and in conditions of a certain sort), then when he exercises that ability/power (in those conditions), he produces the relevant outcome at a rate exceeding chance.

The basic intuition here is that abilities and powers are understood relative to a baseline of chance. Being unreliable obviously differs from being unable and, on any plausible way of approximating the chance rate, there is going to be a margin between chance rates of success and succeeding reliably. You are *enabled* or *empowered* to produce an outcome to the extent that your prospect of producing it exceeds chance. If you succeed at a rate no better than chance, then you lack the relevant ability or power. And if you succeed at a rate worse than chance, then you are *disabled* or *enfeebled*: you're better off just trusting to luck than relying on your own efforts. Approaching matters from a slightly different angle, when relying on luck is your best strategy, you are *helpless*. Empowerment is the antithesis of helplessness. To the extent that you are enabled or empowered, your helplessness diminishes.

Applied to cognition, this theory of powers yields the following view:

> If a person possesses a cognitive ability/power to detect the truth (of a certain sort when in certain conditions), then when she exercises that ability and forms a (thin) belief (on relevant matters and in relevant conditions), she will form a(n approximately) *true belief* at a rate exceeding chance.

This leads to the third and final amendment I propose to the definition we began with: knowledge is approximately true thin belief manifesting *cognitive ability*. I endorse this theory of knowledge, which I call *abilism*.

The evidence that we have cognitive abilities in this sense is utterly conclusive. Innumerable findings in cognitive science demonstrate that humans can form true beliefs at rates exceeding chance. But we needn't consult the science journals to convince ourselves of this point. Innumerable everyday tasks, which we routinely and easily complete, require us to have abilities in the sense just defined. I definitely would not have written this chapter unless I formed true beliefs at rates exceeding chance; you would not be reading this chapter unless you formed true beliefs at rates exceeding chance; more generally, civilization would not exist unless humans generally formed true beliefs at rates exceeding chance. Given the evidence, it is not possible for us to lack the cognitive abilities presupposed by (a non-skeptical version of) abilism.

Knowledge and Luck

GETTIER CASES

Abilism and virtue epistemology solve the Gettier problem in exactly the same way. In each case, the diagnosis is that the Gettier subject has a true belief, but the true belief does not manifest the subject's intellectual virtue or cognitive ability.

This solution presupposes that a Gettier subject does not know. Many philosophers have long maintained that this verdict is completely obvious. But some philosophers have dissented (Sartwell 1992; Hetherington 2013; and, in a different way, Turri 2012b). Moreover, recent work has suggested that the dissident minority represents the ordinary way of understanding Gettier cases (Starmans & Friedman 2012). In particular, people tend to attribute knowledge when a "Gettier subject" initially detected a truth-maker for her belief, even if the explanation for why her belief remains true changes in dramatic and unnoticed ways. For example, suppose that Katie is in her locked apartment. She places a blue pen on her coffee table and goes to take a quick shower. While she is in the shower, a thief breaks into her apartment, takes the original blue pen from the table, and replaces it with another blue pen. Katie is unaware that this happened. When Katie emerges from the shower, does she know that there is a blue pen on her coffee table? The tendency is for people to attribute knowledge to Katie. The basic finding has been replicated multiple times in followup work (Turri 2013a; Nagel, San Juan & Mar 2013, in light of Starmans & Friedman 2013). Is this a problem for abilism?

It needn't be a problem, for three reasons. Each of these reasons is supported by theoretical and empirical findings. First, perhaps it's easier than philosophers have supposed for a true belief to manifest cognitive ability. That is, it could be that the definition of knowledge is correct but philosophers incorrectly applied it to these cases. Second, and relatedly, the "Gettier" cases in question differ importantly from Gettier's original cases, so what is obviously true about Gettier's originals might not be obviously true, or true at all, of the cases in question (as Starmans and Friedman 2013 explain; see also Turri, Buckwalter & Blouw, 2014). Indeed, philosophers have been shockingly careless in continually lumping cases with radically different structures under the heading of "Gettier case" (Blouw, Buckwalter & Turri, forthcoming). Third, these cases inevitably involve complex explanatory relations, so perhaps people's knowledge attributions signify a failure to process important details, at least to some extent.

This third point requires some elaboration. As mentioned above, we can think of Gettier cases as having a tripartite structure: start with a belief sufficiently well formed that it would ordinarily qualify as knowledge, then add an element of bad luck that would ordinarily prevent the well-formed belief from being true, followed by a dose of good luck that cancels out the bad. In order to help people appreciate the underlying tripartite structure of Gettier cases, I conducted an experiment in which one group of people read a story presented in three stages, which thematized the underlying tripartite structure, while a control group read the same case presented all at once instead of in three stages. Whereas people in the control group attributed knowledge roughly half the time (48%), people who read the tripartitioned story attributed knowledge significantly less frequently (29%; Turri 2013a). This is some

evidence that virtue epistemologists have correctly identified the relevant causal structure of Gettier cases that inhibits knowledge.

FAKE BARN CASES

A common objection to the standard virtue-theoretic definition of knowledge is that it faces a class of obvious and stubborn counterexamples known as "fake barn" cases (Goldman 1976 introduced the case into the literature; it is credited to Carl Ginet). In a fake barn case, the agent ("Henry" in the original) perceptually detects an object (a barn in the original), and this perceptual relation remains intact throughout. It is then revealed that the agent currently inhabits an environment filled with numerous fakes ("papier-mâché facsimiles of barns" in the original), which look just like real ones. Some philosophers claim that we are "strongly inclined" to deny knowledge in such cases (Goldman 1976; see also Chisholm 1989: 93; Pritchard 2005; 2014: chap. 6). But because the agent forms a true belief by exercising his highly reliable perceptual faculties, virtue epistemology and abilism entail that the agent has knowledge, which is interpreted as a straightforward counterexample to the view (Pritchard 2009, 2012).

Virtue epistemologists have addressed such cases at length, by either trying to show why their view does not deliver a verdict of knowledge (Sosa 1991: 238–9; Greco 2010: 76ff; see also Jarvis 2013), or developing explanations for why a verdict of knowledge might seem counterintuitive despite being correct (Sosa 2007: lectures 2 and 5). But the correct response is to simply point out that knowledge is present in such cases and that there is nothing counterintuitive about this verdict (Turri 2011a). It is intuitively correct and perfectly natural to attribute knowledge in such cases. The critic's intuitive judgment is idiosyncratic and mistaken.

This is not mere speculation or an intuitional stalemate. Intuitions about fake barn cases have been tested directly and people tend to attribute knowledge (Colaco et al., forthcoming). More generally, in cases where an agent perceives something that makes her belief true, and this perceptual relation remains intact, people overwhelmingly attribute knowledge, even when look-alikes pose a salient threat to the truth of the agent's belief (Turri, Buckwalter & Blouw, 2014). Indeed, common sense basically views these as *paradigm instances* of knowledge, with rates of knowledge attribution often exceeding 80%.

Abilism and virtue epistemology both predict that agents in fake barn cases know, which turns out to be the correct verdict. From a certain perspective, this can seem like a bold prediction whose surprising truth redounds greatly to the theory's credit. But to claim such credit would be cheating, I think, because the perspective in question is warped through overexposure to discussion of fake barn cases in contemporary epistemology. Claiming credit here grants far too much weight to a perspective constituted by idiosyncratic intuitions about

fanciful cases. We make no progress in inquiry by catering to such intuitions. They who blind themselves by walking willfully into the darkness become no more perceptive by stumbling back into the light.

Commonsense Metaphysics and Causal Cognition

Standard virtue epistemology and abilism both define knowledge as a certain sort of representational state (belief or thin belief) that manifests a relevant feature of the knower (reliable virtue or cognitive ability). Some commentators have wondered whether the *manifestation* relation is clear enough to bear the theoretical load assigned to it (Levin 2004) or, worse, whether it is just superfluous jargon (Jarvis 2013: 549n4). As far as I can tell, neither charge withstands scrutiny.

But let me first say that I attach no importance to the name 'manifestation'. It is a convenient label for an certain intimate relation that obtains between powers and outcomes, and I'm happy to relinquish it if comes across as jargon that interferes with effective communication or theorizing. In my own experience, there seems to be widespread agreement on when the relation is present and when it isn't. Indeed, I suspect that the competent use of transitive verbs in natural language presupposes skill at recognizing the relation.

Abstractly put, my view is that, for at least many transitive verbs, V, there is a difference between V-ing something and merely causing it to be V-ed. The former entails the latter, but not vice versa. V-ing something consists in the relevant outcome manifesting the agent's relevant power, whereas merely causing it to be V-ed does not. To take a specific example, consider the verb 'break'. Breaking something is one thing, causing something to be broke another. The former entails the latter, but not vice versa. If I smash the vase into the wall, I break it; it's shattering manifests my physical strength. By contrast, if I, physically incapacitated, pay you to smash it into the wall, you break it, not me, even though I caused it to be broken. Similar remarks apply to countless other verbs, such as 'trap', 'paint', 'count', 'heat', 'lift', etc.

I submit that 'know' is just another verb that works this way (a point anticipated in Turri 2011a, 2011b, 2012b). 'Know' is the most general verb we have for picking out a cognitive relation of interest to us, namely, having one's cognitive abilities manifest in the acquisition of (approximately) true thin belief (compare Williamson 2000: chap. 1). Verbs naming more specific cognitive relations of this sort include 'detect', 'discover' and 'remember'. A correct theory of knowledge features the manifestation relation because the correct theory of the relation picked out by 'know' implicates the manifestation relation. The fact that 'know' fits this pattern makes the inclusion of *manifestation* in a theory of knowledge informative and non-superfluous.

Personal experience, reflection, and social observation make me confident of my claims here about the use of transitive verbs and the intuitive relation of manifestation. But more is, and should be, required to substantiate my claims to the community of researchers working on these questions. To that end, I conducted a preliminary experiment to test whether people reliably apply verbs in the way my view predicts.

The experiment focused on ordinary non-epistemic verbs and the results are encouraging. In a between-subjects design, people read a simple story about an artifact, either a microwave or a trap. The microwave story was based on thought experiments I had proposed in earlier work (e.g., Turri 2012b: 256). It begins with a glass of water sitting in a microwave. The "normal" and "abnormal" versions of the story then continue differently.

(Normal) The microwave starts and draws electricity through its power cord. The microwave sends microwaves that enter the glass and penetrate the water molecules. Soon the water is very hot.

(Abnormal) The microwave short-circuits and the electrical outlet starts on fire. The fire spreads quickly through the kitchen and engulfs the glass of water. Soon the water is very hot.

People were then asked which option better described what happened in the story:

The microwave heated the water.

The microwave merely caused the water to get hot.

The first option uses the transitive verb 'heat' to describe the microwave's relation to the water. The second option uses a derivative causative construction instead.

On my view, the microwave heats the water in the normal case but not in the abnormal case. Moreover, on my view, the correct application of 'heat' implicates recognition of the manifestation relation. So I predict that people will tend to select the first option for the normal story; whereas they will tend to select the second option for the abnormal story. And this is exactly what happened: 71% of people selected the first option for the normal story, compared to 14% for the abnormal story. A very similar pattern of response emerged for the normal and abnormal versions of a different cover story about a cage for trapping rats: 75% of people said that the cage "trapped" a rat in the normal story, compared to 11% for the abnormal story. [1]

The work on Gettier cases discussed earlier provides evidence that people apply 'know' in the way my theory predicts. Indeed, I view Gettier's original cases as the epistemic analogues of the "abnormal" artifact cases mentioned above. While further investigation into the matter is required and ongoing, I suspect it's no coincidence that typically around 10% to 15% of people

attribute knowledge in cases like Gettier's originals, whereas about 80% attribute knowledge in closely matched control cases. This is remarkably similar to the rates at which people used, or withheld use, of the non-epistemic transitive verbs in the experiment reported above. Overall, the current evidence supports my view that 'know' is the verb we use to mark the presence of the familiar manifestation relation for cognitive performances.

More importantly, this all points to an intriguing possibility related to recent work by Fiery Cushman and Liane Young (2011). Cushman and Young provided evidence that representations, distinctions, and patterns exhibited in moral judgment derive from general non-moral psychological mechanisms, including processes of causal cognition. My view is that something similar is true for epistemological judgment—namely, that it inherits some of its most important properties from general non-epistemological psychological mechanisms and representations, such as *manifestation* and *explanatory nondeviance*. Questions in this area will remain at the top of my research agenda moving forward.

Conclusion

My current view is that the theory of knowledge proposed here, abilism, captures the ordinary concept of knowledge and is the clearest, most accurate, and informative philosophical definition of knowledge available. But even if this is correct, it certainly does not mark the end of inquiry into knowledge, because abilism is also a thoroughly naturalistic theory. It identifies in broad outline a real, objective relation suited for scientific investigation: powers producing outcomes. Further insight into the nature of knowledge and knowledge-attributions will come from theoretically informed cognitive science. If I am right, then something akin to Quine's (1969, 1990) vision for epistemology is upon us, though for reasons quite unlike those Quine identified.[1]

References

Alfano, M. (2011). Expanding the situationist challenge to responsibilist virtue epistemology. *Philosophical Quarterly*, *62*(247), 223–249.

Austin, J. L. (1956). A plea for excuses. *Proceedings of the Aristotelian Society*, *57*, 1–30.

[1] Thanks to Christian Miller and Angelo Turri for helpful feedback. This research was supported by the Social Sciences and Humanities Research Council of Canada, the Character Project at Wake Forest University and the John Templeton Foundation (neither of which necessarily endorses any views expressed here), and an Early Researcher Award from the Ontario Ministry of Economic Development and Innovation.

Baehr, J. S. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. Oxford: Oxford University Press.

Blouw, P., Buckwalter, W., & Turri, J. (Forthcoming). Gettier cases: A taxonomy. In R. Borges, C. de Almeida, & P. Klein (Eds.), *Explaining knowledge: New essays on the Gettier problem*. Oxford: Oxford University Press.

Buckwalter, W. (Under review). Factive verbs and protagonist projection. Waterloo, Onatrio: University of Waterloo.

Buckwalter, W., Rose, D., & Turri, J. (Forthcoming). Belief through thick and thin. *Noûs*. Published online on Sept. 15, 2013. doi:10.1111/nous.12048.

Buckwalter, W., & Turri, J. (Under review a). Action, truth and knowledge. Waterloo, Ontario: University of Waterloo.

Buckwalter, W., & Turri, J. (Under review b). In the thick of moral motivation. Waterloo, Ontario: University of Waterloo.

Chisholm, R. (1989). *Theory of knowledge* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Church, I. M. (2013). Manifest failure failure: The gettier problem revived. *Philosophia*, *41*(1), 171–177. doi:10.1007/s11406-013-9418-5.

Colaco, D., Buckwalter, W., Stich, S., & Machery, E. (Forthcoming). Epistemic intuitions in fake-barn thought experiments. *Episteme*.

Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*(6), 1052–1075. doi:10.1111/j.1551-6709.2010.01167.x.

Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*(1), 100. doi:10.1037/h0034449.

DeRose, K. (1996). Knowledge, assertion and lotteries. *Australasian Journal of Philosophy*, *74*(4), 568–580.

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Noûs*, *32*(4), 504–530.

Doris, J. M. (2002). *Lack of character: personality and moral behavior*. Cambridge: Cambridge University Press.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, Mass.: MIT Press.

Flanagan, O. (1991). *Varieties of moral personality*. Cambridge, MA: Harvard University Press.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, *73*(20), 771–791.

Greco, J. (1993). Virtues and vices of virtue epistemology. *Canadian Journal of Philosophy, 23*(3), 413–432.

Greco, J. (2003). Knowledge as credit for true belief. In M. DePaul & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology* (pp. 111–134). New York: Oxford University Press.

Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge: Cambridge University Press.

Greco, J. & Turri, J. (2011). Virtue epistemology. *The Stanford Encyclopedia of Philosophy* (Winter 2013 ed.). Ed. E. N. Zalta. http://plato.stanford.edu/archives/win2013/entries/epistemology-virtue/.

Greco, J., & Turri, J. (2012). Introduction. In *Virtue epistemology: contemporary readings*. Cambridge, MA: MIT Press.

Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society*, 99, 315–331.

Hartshorne, H., & May, M. A. (1928). *Studies in deceit. Book I: General methods and results. Book II: Statistical methods and results*. New York: Macmillan.

Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research*, *80*(3), 497–522.

Hazlett, A. (2012). Factive Presupposition and the Truth Condition on Knowledge. *Acta Analytica*, *27*(4), 461–478.

Hetherington, S. (2013). There can be lucky knowledge. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology*. (2nd ed.). Malden, Mass.: Wiley-Blackwell. 164–76.

Jarvis, B. (2013). Knowledge, Cognitive Achievement, and Environmental Luck. *Pacific Philosophical Quarterly*, *94*, 529–551.

Kallestrup, J., & Pritchard, D. (2014). Virtue Epistemology and Epistemic Twin Earth. *European Journal of Philosophy, 22*(3): 335–357. First publication online, Dec. 2011. doi:10.1111/j.1468-0378.2011.00495.x.

Kvanvig, J. L. (1992). *The intellectual virtues and the life of the mind*. Lanham, MD: Rowman and Littlefield.

Levin, M. (2004). Virtue epistemology: No new cures. *Philosophy and Phenomenological Research*, *69*(2), 397–410.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper Perennial.

Miller, C. (2014). Moral virtues, epistemic virtues, and the Big Five. In O. Flanagan & A, Fairweather (Eds.), *Naturalizing Epistemic Virtue* (pp. 92–117). New York: Cambridge University Press.

Mischel, W., & Peake, P. K. (1982). Beyond déjà vu in the search for cross-situational consistency. *Psychological Review*, *89*(6), 730–755.

Murphy, M. C. (1998). Review of Virtues of the Mind. *Philosophical Books 39*(3): 210–212.

Murray, D., Sytsma, J., & Livengood, J. (Forthcoming). God knows (but does God believe?). *Philosophical Studies*. doi:10.1007/s11098-012-0022-5.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that P without believing that P. *Noûs 47*(2), 371–384.

Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true beliefs. *Cognition*, *129*(3): 652–661. doi:10.1016/j.cognition.2013.02.008.

Olin, L., & Doris, J. M. (2014). Vicious minds. *Philosophical Studies*, *168*(3), 665–692 doi:10.1007/s11098-013-0153-3.

Pritchard, D. (2005). *Epistemic luck*. New York: Oxford University Press.

Pritchard, D. (2009). Apt performance and epistemic value. *Philosophical Studies*, *143*(3), 407–416.

Pritchard, D. (2012). Anti-luck virtue epistemology. *Journal of Philosophy*, *109*(3), 247–279.

Pritchard, D. (2014). *What is this thing called knowledge?* (3rd ed.). New York: Routledge.

Pritchard, D. & Turri, J. (2011). The value of knowledge. *The Stanford Encyclopedia of Philosophy* (Winter 2012 ed.). Ed. E.N. Zalta. http://plato.stanford.edu/archives/win2012/entries/knowledge-value/.

Quine, W. V. (1969). Epistemology naturalized. In *Ontological relativity and other essays*. New York: Columbia University Press.

Quine, W. V. (1990). Norms and aims. In *The pursuit of truth*. Cambridge, MA: Harvard University Press.

Radford, C. (1966). Knowledge—by examples. *Analysis*, *27*(1), 1–11.

Riggs, W. D. (2002). Reliability and the value of knowledge. *Philosophy and Phenomenological Research*, *64*(1), 79–96.

Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford: Oxford University Press.

Rose, D., Buckwalter, W., & Turri, J. (2014). When words speak louder than actions: delusion, belief and the power of assertion. *Australasian Journal of Philosophy*. doi:10.1080/00048402.2014.909859

Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*. London: Pinter & Martin.

Sartwell, C. (1992). Why knowledge is merely true belief. *The Journal of Philosophy*, *89*(4), 167–80.

Sosa, E. (1991). *Knowledge in Perspective*. Cambridge: Cambridge University Press.

Sosa, E. (2003). The place of truth in epistemology. In M. DePaul & L. Zagzebski (Eds.), *Intellectual virtue: Perspectives from ethics and epistemology* (pp. 155–79). Oxford: Oxford University Press

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge*, vol. I. Oxford: Oxford University Press.

Sosa, E. (2011). *Knowing full well*. Princeton, NJ: Princeton University Press.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, *124*(3), 272–283.

Starmans, C., & Friedman, O. (2013). Taking "know" for an answer: A reply to Nagel, San Juan, and Mar. *Cognition*, *129*(3), 662–665.

Stroud, B. (1984). *The significance of philosophical skepticism*. Oxford, England: Clarendon Press.

Turri, J. (2010). On the relationship between propositional and doxastic justification. *Philosophy and Phenomenological Research*, *80*(2), 312–326.

Turri, J. (2011a). Believing for a Reason. *Erkenntnis*, *74*(3), 383–397. doi:10.1007/s10670-011-9271-5.

Turri, J. (2011b). Manifest failure: The Gettier problem solved. *Philosophers' Imprint*, *11*(8), 1–11.

Turri, J. (2011c). Mythology of the factive. *Logos & Episteme*, *2*(1), 143–152.

Turri, J. (2012a). In Gettier's wake. In S. Hetherington (Ed.), *Epistemology: The key thinkers*, 214–29. London: Continuum.

Turri, J. (2012b). Is knowledge justified true belief? *Synthese*, *184*(3), 247–259.

Turri, J. (2012c). Review of *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, by John Greco. *Mind*, *121*(481), 183–187. doi:10.1093/mind/fzs032.

Turri, J. (2013a). A conspicuous art: Putting Gettier to the test. *Philosophers' Imprint*, 13(10), 1–16.

Turri, J. (2013b). Knowledge guaranteed. *Noûs*, *47*(3), 602–612.

Turri, J. (2014). The problem of ESEE knowledge. *Ergo*.

Turri, J. (Forthcoming a). Knowledge as achievement, more or less. In M. Á. F. Vargas (Ed.), *The present and future of virtue epistemology*.

Turri, J. (Forthcoming b). Epistemic situationism and cognitive ability. In M. Alfano & A. Fairweather (Eds.), *Epistemic situationism*. Oxford: Oxford University Press.

Turri, J. (Forthcoming c). Unreliable knowledge. *Philosophy and Phenomenological Research*. doi:10.1111/phpr.12064

Turri, J. (Under review). Ordinary unreliable knowledge. University of Waterloo.

Turri, J., Buckwalter, W., & Blouw, P. (2014). Knowledge and luck. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0683-5.

Turri, J., & Friedman, O. (2014). Winners and losers in the folk epistemology of lotteries. In J. R. Beebe (Ed.), *Advances in experimental epistemology*, 45–70. London: Continuum.

Turri, J., Rose, D., & Buckwalter, W. (Under review). Choosing and refusing: Doxastic voluntarism and folk psychology. University of Waterloo.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

Zagzebski, L. (1994). The inescapability of Gettier problems. *Philosophical Quarterly*, *44*(174), 65–73.

Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.

Zagzebski, L. (2009). *On epistemology*. Belmont, CA: Wadsworth.

PART } VI

Particular Virtues

15 }

Christian Humility as a Social Virtue

Michael W. Austin

In this chapter, I explain and defend an important but underappreciated aspect of the Christian virtue of humility. I take Christian humility to have both self-regarding and other-regarding elements. Many analyses of humility, both Christian and secular, emphasize its self-regarding aspects. Such accounts of humility focus on self-knowledge as the primary aspect of this virtue, with an emphasis on knowledge of one's limits, dependence upon others, and a related tendency to refrain from exaggerating one's talents, gifts, and abilities. The humble person gives credit where credit is due for these goods. For the Christian, credit will ultimately go to God; whereas for the philosophical naturalist it will go to no one (or perhaps to blind chance).[1] These self-regarding aspects of humility are related to its other-regarding aspects. Exaggeration of one's talents can lead to an exaggeration of one's self-worth, which may have an impact on how one relates to others. For example, such mental attitudes may cause or reinforce selfish behavior.

In what follows, I focus on the other-regarding aspects of Christian humility. Or, to put it differently, I focus on Christian humility as a social virtue. Humility, on a Christian view, is an important virtue for how we relate to God and other human beings. My primary aim is to offer an analysis of the social aspects of this virtue as it concerns interpersonal relationships with other people. In what follows, I first examine two key New Testament passages that deal with humility as a social virtue. Next, I describe an assumption present in this chapter concerning the modularity of virtue. I then discuss three important claims about the nature of social humility and the ideally humble Christian person: (i) the humble person believes that he ought to have a *prima facie* preference for the

[1] This is suggested by Wielenberg (2005): 102–112. For another non-religious account along these lines, see Richards (1992). For Christian accounts that emphasize the self-regarding aspects of humility, see Taliaferro (1989); and Copan (2006).

satisfaction of the interests of others over the satisfaction of his own interests; (ii) the humble person has a *prima facie* preference for the satisfaction of the interests of others over his own; and (iii) the humble person will be disposed to engage in self-sacrificial actions for the good of others. Finally, I conclude by considering two objections to my account of Christian humility, in order to offer both a limited defense and further explanation of its contents.

Social Humility in the New Testament: Two Important Passages

The idea that humility is in part a social virtue is reflected in several passages from the New Testament.[2] Before analyzing the virtue in more detail, I will first briefly examine two passages.

The Greek term *tapeinophrosune* is translated as "humility" in two passages that are important for understanding this virtue, namely, 1 Peter 5:5–6 and Philippians 2:1–11.[3] This Greek term can also be translated as "lowliness of mind."[4] Without examining the particular context in which the term appears in the scriptures, its meaning remains somewhat vague. It is clear that in some contexts the meaning will include a self-regarding aspect, insofar as *tapeinophrosune* is concerned with particular types of self-knowledge. However, it is also clear that this trait possesses other-regarding elements. Verlyn Verbrugge interprets *tapeinophrosune* and most of its cognates in both the Hebrew Bible and New Testament as describing something that is needed for a proper relationship with God and other human beings.[5] In this chapter, the emphasis will be on human relationships, rather than the divine-human relationship. That is, the focus in what follows will be on some of the interpersonal elements of humility, on how this virtue should operate within relationships between human beings. This will help us to understand how humility functions as a social virtue. With this in mind, I will now briefly examine the aforementioned New Testament passages with a special focus on the social aspects of the Christian virtue of humility.

The New Testament epistle of 1 Peter was intended to be circulated among several early Christian communities.[6] For a discussion of humility, elements of 1 Peter 5 are particularly relevant. In this chapter, the author first gives instructions to those in positions of church leadership and exhorts them to carry

[2] For more on the significance of taking humility to be a social virtue in the context of the ancient Near East, see Dickson and Rosner (2004).

[3] Two other places where this term appears are Ephesians 4:2 and Colossians 3:12. Significantly, the context in both of these passages also points to humility as a social virtue, as it appears in a discussion of relationships and other virtues appropriate for them, such as compassion, kindness, forgiveness, and patience.

[4] Verbrugge (2000): 555.

[5] Ibid.

[6] See 1 Peter 1:1–2.

out their responsibilities with integrity. The writer then encourages those who are under the care of these leaders to submit themselves to their leadership, and goes on to instruct the epistle's readers as follows: "Clothe yourselves, all of you, with humility toward one another, for 'God opposes the proud but gives grace to the humble.' "[7] While no definition of humility is given here, it is assumed that humility is at least in part a virtue that the Christian exemplifies in relation to other human beings; it is a virtue that Christians are to have "toward one another." It is also a virtue that Christians can take on, as they might take on a piece of clothing. It is a virtue that can be intentionally pursued and exemplified. I will expand on these points as I analyze humility below. First, in order to obtain more clarity about the nature of humility as a social virtue in the New Testament, I will turn to a crucial passage for understanding this virtue as it was thought to be exemplified by Christ.

In his letter to the Philippians, Paul first expresses gratitude for their partnership with him and then exhorts them to work out their salvation in unity by seeking to live in a manner worthy of the gospel of Christ. Next, in Philippians 2:1–11, Paul offers important theological and ethical information related to the humility of Christ and its ramifications for Christian thought and practice:[8]

> So if there is any encouragement in Christ, any comfort from love, any participation in the Spirit, any affection and sympathy, complete my joy by being of the same mind, having the same love, being in full accord and of one mind. Do nothing from selfish ambition or conceit, but in humility count others more significant than yourselves. Let each of you look not only to his own interests, but also to the interests of others. Have this mind among yourselves, which is yours in Christ Jesus, who, though he was in the form of God, did not count equality with God a thing to be grasped, but emptied himself, by taking the form of a servant, being born in the likeness of men. And being found in human form, he humbled himself by becoming obedient to the point of death, even death on a cross. Therefore God has highly exalted him and bestowed on him the name that is above every name, so that at the name of Jesus every knee should bow, in heaven and on earth and under the earth, and every tongue confess that Jesus Christ is Lord, to the glory of God the Father. (Phil. 2:1–11 ESV)

---

[7] 1 Peter 5:5 (English Standard Version). In the next verse, the author goes on to discuss humility in relation to God, saying "Humble yourselves, therefore, under the mighty hand of God so that at the proper time he may exalt you."

[8] Strimple (1979). Many New Testament scholars take verses 5–11 to be an early Christian hymn, which Paul is appropriating in his letter to the Philippians. Nevertheless, I follow Strimple, who maintains that as we study the passage, we must see the hymn in its literary context in the epistle, and interpret it accordingly. That is, it is important that we interpret Philippians 2:5–11 in a way that gives priority to its role in Paul's argument in the epistle, rather than isolate the hymn from this rhetorical context.

Paul exhorts the Philippians to be united for the sake of the gospel and to imitate the humility exemplified by Christ through his obedience to God the Father in the incarnation and crucifixion. Though much theological reflection on this passage centers on interpretations of the controversial concept of Christ's *kenosis* (i.e., self-emptying), the passage itself is also (and perhaps primarily) ethical in nature.[9]

One important focus of the passage, in verses 6–8, is the humility of Jesus.[10] In his discussion of these verses, Larry Hurtado claims that the "self-humbling and obedience in verse 8 are probably to be taken as referring to the readiness of the human Jesus to take the path of service (and this must be taken as obedience to God), even at the cost of his life through the violence of crucifixion."[11] The passage claims that Jesus opted to become incarnate and suffer the pain and humiliation associated with crucifixion. In addition, it is significant that Paul instructs the Philippians to imitate the humility of Christ in their particular context.

Given that Christ's humility as it is exemplified in the Christian doctrines of the incarnation and crucifixion provides the theological foundation for the ethical exhortation present in the passage, a main thrust of the passage has to do with humility and its connections to unity within this early Christian community. That is, the passage is focused on humility as a social virtue, a virtue that is important for interpersonal human relationships. Humility, then, is not merely a self-regarding virtue, but an other-regarding one as well. Humility turns its possessor away from herself, and toward the needs, interests, and intrinsic value of others. As C. S. Lewis aptly (but perhaps a bit too strongly) puts it, "[A] really humble man . . . will not be thinking about humility: he will not be thinking about himself at all."[12] In order to analyze Christian humility as a social virtue, I assume that there are modules of virtue. In the next section, I discuss what this means.

The Modularity of Virtue

In what follows I assume that there are modules of virtue. According to Robert Adams, modules of the virtues are behavioral dispositions that are independent and domain-specific.[13] As he puts it, "[A] person will often acquire and exercise a disposition to act in a certain way in one domain without being disposed to

[9] For an analysis of the metaphysics of the *kenosis*, see DeWeese (2007).
[10] Hurtado (2005): 83–107.
[11] Ibid.: 103.
[12] Lewis (1952): 99.
[13] Adams (2006): 25–27.

act similarly in somewhat different domains."[14] A person might be disposed to be courageous in domain X (sports, for example), but not so disposed in domain Y (speaking up for his moral convictions in the face of opposition). Modules of a particular virtue can be added together to form a more complete disposition, a more wide-ranging virtue. On this view, the more courageous one is, the more domains there will be in which one is disposed to behave in a courageous manner. It is also the case that one is more courageous as the depth of the relevant dispositions increases within these domains. These modules, then, can be thought of as parts of a virtue, and are morally valuable in their own right.

Adams focuses on modules as independent and domain-specific components of a virtue. Two points must be added to clarify how I conceive of such modules. First, one particular module of a virtue might obtain in a variety of domains. That is, a single discrete component of a virtue can surface in a variety of contexts. For example, one might have a physical module of courage across more than one domain. A police officer may show physical courage in that context when pursuing a suspect, and also in the context of sports. Second, I believe that we can conceive of modules of virtue, not just in terms of the domains across which a behavioral disposition may exist, but also in terms of the relevant cognitive, emotional, and active aspects (i.e., the actions the trait disposes one to perform) of a particular trait. On this modification of Adams's view that is assumed in what follows, modules are not only distinguished by separate domains, but by a virtue's cognitive, emotional, and active aspects as well. These aspects of humility as described below can be thought of as modules of the complete form of Christian humility that are expressible in a variety of contexts.

In what follows, I examine one module from each of the following categories: cognitive, emotive, and active. I take the cognitive modules of humility to include belief and knowledge. For my purposes, emotive modules include emotions as well as certain types of desires and preferences, which will be clarified later in the chapter. It is also important to note that the boundaries between these categories of modules are sometimes porous. For instance, emotions may also be cognitive, in certain respects.[15] Finally, the active modules concern particular dispositions to act that the humble person will possess, which yield certain types of acts that are characteristic of her. With this in mind, I now discuss three significant social modules of humility as exemplified by the ideally humble Christian human person.[16]

---

[14] Ibid.: 125.

[15] Emotions arguably involve judgments. See Nussbaum (2001).

[16] The account is not intended to be comprehensive. For example, I do not discuss other modules of humility which I take to also be constitutive of this moral virtue. I focus on the other-regarding aspects of humility, rather than the more personal self-regarding aspects, though these are connected to one another. Nor do I discuss intellectual humility in what follows, though a full account must include this aspect of the virtue. For more on intellectual humility, see Roberts and Wood (2007): 236–256; and Pardue (2013). See also Fullam (2009) for a theological account of humility based on Aquinas that is in some ways similar to the account I offer here.

(CH) The humble person believes that he ought to have a *prima facie* preference for the satisfaction of the interests of others over the satisfaction of his own interests

(CH) is a cognitive module of humility which involves intentionally taking a certain perspective about one's obligations to others relative to oneself. This shift in our thinking is necessary on a Christian anthropology, given the view that we possess a propensity for self-centeredness. This is rational because the Christian can entrust his welfare to God, rather than seeking to manage it himself, which yields a particular freedom to focus more heavily on the welfare of others. Because of this, leaders and followers can be humble toward one another, as 1 Peter 5 describes. Moreover, in a community that functions in a way that is consistent with Paul's instructions in the letter to the Philippians, the Christian also entrusts his welfare to other members of his community as a means by which God cares for him. This point is stated well by John Whittaker:

> Those who believe in the promises of God are unbound from the task of achieving their happiness. They have replaced the struggle for happiness with the trust that it will be given to them in God's way. Thus, they are free to show others a pale copy of the concern that they believe has been shown to [them].[17]

This trust in God for one's happiness, which can enable one to show concern for others, is present in Philippians 2. For Paul, if members of the Philippian church follow his instructions, they will experience something like the care of God alluded to by Whittaker via one another. God's care is manifested in their relationships with each other. But for this to happen, humility must be a characteristic of individual members in the community. Paul morally exhorts the Philippians to intentionally take a particular perspective regarding others in their community, "in humility" they are to "count others more significant" than themselves. They are to see others in a particular way—namely, as being "more significant" than themselves.

What might this mean? On some interpretations of seeing others as being more significant than oneself, potential moral and theological problems arise. Morally, it seems wrong to take some human beings as having greater worth or significance than others. There are strong biblical and theological reasons in support of the view that all human beings are equal with respect to their intrinsic worth and basic dignity, grounded in their status as creatures made in and reflective of the image of God. Theologically, if Christ is taken to be the exemplar of humility, then problems can arise for the claim that he saw others as being more significant than himself. Christians ought to see themselves as

[17] Whittaker (1992): 238.

being less important than Christ. Moreover, if Christ genuinely saw a mere human to be more important or valuable than himself, then, given traditional claims about his divinity, this would seem to be a case of self-deception. This, of course, would be an unsatisfactory result.

There is another theological problem looming here related to ascribing such humility to God the Father. Taking humility to require that others are more significant (i.e., valuable) than oneself would then have the undesirable result that God the Father sees his creatures as more significant than himself. Given the predominant understandings of God in Christian theology, this would also be a serious mistake. The Anselmian God, for instance, should not see human creatures as having more intrinsic worth than he, given the properties ascribed to him.

These problems, however, can be avoided. In Philippians 2:1–11, which is the primary biblical and theological source for (CH), Paul clarifies what is meant by taking others to be more significant than oneself. First, he contrasts taking others to be more significant than oneself with performing actions that are selfish. This sheds light on the intended meaning of seeing others as being more significant than oneself, insofar as the contrast points to actions and motives for actions that are selfish and done from a vain conceit, rather than an evaluation of intrinsic worth. The humble person is unselfish. The claim that the Philippians are to consider others as more significant than themselves is also clarified in context by Paul's exhortation to the Philippians that each of them should "look not only to his own interests, but also to the interests of others." It seems to me that the best way of understanding what it means to see others as being more significant than oneself in light of the context of the passage in particular concerns *how we weigh the interests of others relative to our own*. More specifically, we are to *prefer* the satisfaction of their interests to the satisfaction of our own interests. (CH), then, does not include the belief that the interests of others are objectively more important or valuable than my own. Instead, it means that I believe that I ought to prefer, *prima facie*, that their interests be satisfied over my own. The content of the belief need not be understood or stated by the belief holder in these precise terms, however. The humble person may simply believe, "I should usually put others before myself." A fuller philosophical analysis, however, will spell this out in a more precise manner.

The foregoing interpretation of what it means to count others as being more significant than oneself is supported by the description in the passage of Christ's incarnation and crucifixion as actions indicative of his humility which also exemplify a preference for satisfying the interests of others. The ideally humble person believes that she ought to have such a preference as a follower of Christ seeking to imitate him. Christ exemplifies this preference for satisfying the interests of others, and this serves to justify (in part) the belief alluded to in (CH) for those who seek to imitate his example in some analogous way.

The preference referred to in (CH) is a *prima facie* preference, as circumstances and context will dictate when one may prefer the satisfaction of her own interests. I offer a fuller analysis of this preference, including the *prima facie* condition, in the next section. The central point at present is that the ideally humble person will accept as true the claim that she ought to have this preference.

(EH) The humble person has a *prima facie* preference for the satisfaction of the interests of others over his own

Given that it goes beyond mere belief in important ways, I have chosen to categorize (EH) as an emotive module of the virtue of humility.[18] There is a vast philosophical literature dealing with the nature of belief, emotion, desire, and preference, a discussion of which would take us too far afield. In what follows, I will make use of portions of the literature, without seeking to comprehensively discuss the state of the debate. With this in mind, I now turn to a discussion of (EH).

The readers of 1 Peter are encouraged to be humble toward one another. To understand what this entails, it will be helpful to again examine the ultimate Christian exemplar. In Philippians 2, Christ is pictured as exemplifying (EH). In the voluntary restriction of his divine capacities involved in the incarnation and crucifixion, motivated by a desire to bring about the redemption of human beings as well as the rest of creation,[19] it is clear that Paul takes Christ to have preferred the satisfaction of the interests of others over the satisfaction of his own interests for the sake of the glory of God. Christ's actions on our behalf are conducive to the satisfaction of our interests because, on a picture of human flourishing present in the Christian scriptures and tradition, it is in our interest to know and love God, as well as to love our neighbors as ourselves.[20] In Christian theology, Christ's life, death, and resurrection make it possible for us to receive and participate in these spiritual and relational goods. While both one's own interests and those of others are important (and equally so from God's impartial perspective), one implication of Philippians 2:1–11 for the individual follower of Christ is that she ought to prefer others insofar as she places greater subjective weight, *prima facie*, on their interests compared to her own.

---

[18] The humble person's preference need not include the concept of being *prima facie*, per se, but she will have the preference alluded to by (EH) in perhaps less complicated terms. She possesses this module of humility if she generally prefers others' needs and interests over her own. A deeper and thicker account, however, as I am offering, will include such concepts.

[19] See Romans 8:1–25.

[20] See Matthew 22:34–40. The point is not that we should love others merely because it will make us happy. Rather, it is because that is what the Christian exemplar did, and it is generally conducive to both their flourishing as well as our own.

To understand (EH) in more detail, consider Philip Pettit's discussion of the nature of a preference.[21] A preference is not a mere taste, such as the taste for amber ale or the color red. Tastes are immune from deliberation; whereas preferences are not. Preferences are also distinct from mere desires insofar as a preference includes a ranking of at least two alternatives. We do not prefer *X*; rather, we prefer *X* over *Y*. Mere desires do not possess this comparative aspect. In addition, a preference may be concerned with the self-interest of its possessor, but it may not. One may prefer an alternative that is worse with respect to one's own welfare.

In the case of (EH), imagine an agent has a choice between *A* (satisfying his own interests) and *B* (satisfying the interests of another). Imagine as well that the interests in play are equally valuable regarding their potential contribution to the welfare of both parties; in this way they are equally significant for each individual. In such a case, the ideally humble Christian agent will believe that *B* has a desirable property that *A* lacks, namely, that it fulfills the obligation alluded to by (CH), which is itself partially grounded in her obligation to imitate Christ as her exemplar of humility. This preference will not always hold or inevitably lead to a correlative action, however, given the *prima facie* condition included in (EH). In addition, it is important to emphasize again that the agent prefers satisfying the interests of others, *prima facie*, over her own, not because they or their interests have greater objective value, but rather because of her belief that she ought to imitate Christ—including the specific belief in (CH) discussed earlier—and her desire to engage in this form of imitation of Christ. Such imitation connects to her goal of having a character that is increasingly similar to Christ's. Finally, there may not be a particular phenomenal experience associated with and perhaps supporting the preference for *A* over *B*. However, it is better if such a qualitative experience is present alongside the preference, as in the ideally humble person the virtue of humility will be present all the way down. Such a person will not only believe that she ought to be humble in these ways—she will deeply desire that others flourish. *Ideally*, there will be such an experience. Instead of an internal struggle, the ideally humble person will exemplify internal psychological unity and integrity that includes felt desires for the welfare of others.

An important point in relation to (EH) is that the humble person will need to have a particular type of correlative belief (or judgment) concerning the interests of others.[22] Another person may believe that some putative interest

---

[21] Pettit (2006): 131–153.

[22] (EH) may relate to (CH) in different ways in any particular case. Ideally, (EH) will be in part a product of (CH). In this way, the preference is securely grounded in a belief about one's obligations, such that when the preference is less strong or absent, the agent will then do what she can to cultivate or rehabilitate it, in part because she believes (CH). I discuss methods for forming this preference later in the section.

would be conducive to her own flourishing, and be mistaken about this. The humble person is not bound by the beliefs and judgments of others concerning their own interests. Rather, she depends on her own beliefs and judgments, which may be informed by the views of others. She will attempt to determine whether or not a putative interest is a genuine interest, and this will inform her preference and the decision that she makes in any particular case. Cognitively, if she determines that the putative interest in question is a genuine interest, then the preference will be rational insofar as she reasonably believes that preferring the interests of the other in this case is best (or at least good) for that other.[23] In this way, she satisfies her obligation in these circumstances to imitate the humility of Christ.

How might one go about intentionally forming (EH)? (EH) involves a way of seeing others that is in some ways analogous to what epistemologists call *seeing as* in the realm of perception.[24] When I see something that I take to be a bird, I will often see it as having further properties. I might see a bird as being large, swift, or black. A useful illustration for my present purposes is the duck/rabbit line drawing. Depending on the perspective one intentionally adopts, one can see it as either a duck or a rabbit. The humble person, then, intentionally sees the satisfaction of the interests of others as preferable to the satisfaction of her own interests. This is not a full-fledged belief that the interests of others are objectively more important than her interests.[25] Rather, (EH) can be construed as including *a willful appearance state*. As one may choose to see the duck-rabbit line drawing as either a duck or a rabbit, in an analogous manner one may choose to see the satisfaction of the interests of others as preferable in the way just described. As Roderick Chisholm explains, there is a descriptive use of the word "appear" which is distinct from belief.[26] For example, one may say "This appears white to me" and mean by it that one is having an appearance of whiteness and that the object being perceived is in fact white. However, one may also mean that the object *appears* white in this light, while knowing that *in reality* the object is gray. In the present context, the suggestion is that we can analogously choose to see the satisfaction of the interests of others as preferable compared to the satisfaction of our own interests, as described by (EH). We can *willfully* or *intentionally take* the satisfaction of the interests of others to be *prima facie* preferable to the satisfaction of our own interests, while knowing that in another sense their interests are equal in importance to our own, all else being equal.

---

[23] Piller (2006): 177.

[24] Audi (2011): 26.

[25] This is where the analogy to the duck/rabbit ends. The drawing has objective properties which enable it to be seen as a duck or a rabbit. The point of the analogy is merely to show the power of choice relative to some appearance states.

[26] Chisholm (1977): 26–30.

We can arrive at such a perspective in part by intentionally focusing on both the interests of others and the desirability of their satisfaction. We attend to the interests of others, rather than our own. This is in part how we may place ourselves in the relevant willful appearance state. The interests of others are in the foreground, and our interests fade into the background of our attention. Then, in part because of this willful appearance state which is a result of our intentional focus, we may more easily find ourselves with the preference described in (EH).

Finally, the significance of the individual interests in play is one important factor in determining whether or not an all-things-considered preference for the satisfaction of the interests of the other is appropriate in any particular situation. However, this is not the sole determining factor. That is, the claim in (EH) is not merely that the humble person should have a *prima facie* preference concerning the interests of others only when their interests are equally or more significant than her own. Rather, I want to suggest (but not argue) for the claim that there may be cases in which the humble person sacrifices some of her own interests for the sake of the interests of others, even when the former are, objectively speaking, more significant than the latter. This is part of the radical humility ascribed to Christ in Philippians 2:1–11, and it is an ideal that the follower of Christ may aspire to in her own particular circumstances. In the next section, I examine a disposition to perform a particular type of action that is characteristic of the ideally humble person.

(AH) The humble person will be disposed to engage in self-sacrificial actions for the good of others

There is less to discuss regarding this active module of humility, given that it is in many ways a product of the previous two modules. In order to clarify what is meant by an active module of a virtue, it is important to see that the acts are not themselves identical to the virtue, but rather they are products of it which can nevertheless be characterized by virtue terms. It is not just that an act is performed *from* a particular virtue, but also that the act itself is virtuous when this is the case. An agent may act from the virtue of humility, and the act itself also be characterized as a humble act. The humble person will be disposed to perform humble acts, if they have the virtue. (CH) and (EH) can contribute to the presence of the disposition alluded to by (AH), but there very well may be other beliefs, emotions, and preferences that contribute to (AH). It is important to examine the type of actions yielded by the virtue because my focus is not merely the virtue of Christian humility itself. I am also concerned with what is true of the ideally humble Christian person.

The disposition to act unselfishly for the good of others is an essential aspect of the Christological account of humility. Philippians 2:1–11 includes

the presence and practice of the concepts present in (AH) in its description of the life and death of Jesus aimed at the redemption of all human beings. The depth and significance of the sacrifice of Christ, including but not limited to the crucifixion, is difficult to comprehend. In the passage, Jesus is portrayed as willingly setting aside his divine prerogatives by becoming human in order to redeem human beings. For this reason, he experienced life as a human infant, child, and adult, with its trials and triumphs. He sacrificially subjected himself to ridicule, persecution, violent abuse, and execution by those whom he came to redeem, out of love for them. It is difficult to imagine a clearer case of a set of actions that are both self-sacrificial and done for the good of others, from a Christian perspective. While human beings cannot replicate the *particular* actions ascribed to Christ in this passage, those who seek to imitate Christ may perform actions that are of a similar *type* as described by (AH).

Objections to the Account

In this final section, I consider two objections that could be raised against the foregoing account of humility. The first objection I consider is the claim that two of the foregoing modules of humility, (CH) and (EH), are false. The second objection to which I reply is the claim that both of these modules, as well as (AH), are too demanding and are therefore implausible. My replies to these objections serve not only to defend the foregoing claims about these social aspects of humility, but also to explain the account in more detail.

(CH) AND (EH) ARE FALSE

Recall that (CH) is the claim that the humble person believes that he ought to have a particular preference, namely, (EH), which is the claim that such a person has a *prima facie* preference for the satisfaction of the interests of others over his own interests. In reply to these claims, the critic may argue that the interests of others are not in fact more important than one's own, and so we should neither believe (CH) nor prefer the satisfaction of the interests of others over the satisfaction of our own as indicated by (EH). One reason to hold this is that from an impartial perspective the inherent value and interests of all humans are equal. There is no property from a God's-eye view that one person possesses which makes her or her interests more important than those of anyone else. From the impartial perspective that many hold the Christian God to possess, all human beings are thought to have equal inherent dignity and worth. The objection, then, is that (CH) and (EH) are false because they endorse a form of partiality. If this is right, then they should

not be included as modules of the virtue of humility, because a virtue should not require one to hold a false belief or to possess a misguided preference.

What can be said in response to this?[27] The objection at issue can be dealt with by employing the concept of role obligations. The humble person knows that he has the responsibility to further the interests of others and work for their good, and that this should take precedence over his own interests. Such a person can know that her interests and those of her closest friend have equal weight from God's point of view, yet consistently adopt a perspective in which she takes her friend's interests to be more important than her own, because we possess some obligations by virtue of the particular roles we inhabit in our personal and social lives. In this instance, the obligations in question relate to her role as a friend. Consider a few other examples. Spouses have obligations to one another qua spouses, which they voluntarily assumed when they recited their wedding vows. A parent has particular obligations qua parent; a nurse, qua nurse; a minister, qua minister, and so on. A person possesses such role obligations only when she stands in the relevant role.

Given the scope of Paul's exhortation to humility in his letter to the Philippians, we can take believing (CH) and possessing (EH) to be partially constitutive of the role played by an individual as a *brother* or *sister* within a Christian community, and as one primary aspect of the more general role obligations Christians have as *followers of Christ*. If (CH) and (EH) are connected with role obligations, then this would help to undermine the force of the claim that they are false, given the nature of such obligations. For example, in a marriage relationship, part of what it means to stand in that relationship is that I take the interests of others to have less weight than those of my spouse in my deliberations and actions. This type of partiality does not entail that her interests always trump those of others. For example, I might sacrifice one of her less important interests in order to save the life of a stranger, as an obvious example. The point is that my role as a husband entails that I regard the interests of my wife to be more important than mine and others, all else being equal. Or consider the case of parents and children. While it is true that the interests of all children carry equal weight, the person who is a parent of a particular child in some sense takes that child's interests to be more important than those of all other children, insofar as he devotes his time, energy, and resources to that particular child. He is taking the interests of that child to be more important

[27] James Spiegel (2003) makes a point that is relevant to this objection. He states that "the humble person's 'low self-regard' may be taken in a *performative* sense, such that one *plays the role* of the unworthy or *assumes* the position of the lowly through her behavior, words, or self-conception" (p. 133). Similarly, perhaps the humble person's deference to the interests of others could be thought of in this performative sense, rather than in a cognitive one. However, Spiegel argues, and I agree, that humility includes both cognitive and performative (similar to what I am calling "active") elements. Given this, we will need to look elsewhere for a full solution to the current objection, a solution which makes use of the active elements of humility while retaining the cognitive elements as well.

for him qua parent than the interests of other children because he is *that* child's parent.[28]

In an analogous manner, in the role responsibilities one possesses as a follower of Christ, one is to view others as more important than himself in the sense of prioritizing their interests. When one takes on these roles, others are more important in this sense from the perspective of the role. This is true, even though from God's perspective, all humans are of equal importance. From God's perspective all of our interests carry the same weight, but from the chosen perspective of an individual in his role of follower of Christ the interests of others are to count more than his own. In this sense, then, both (CH) and (EH) are true.

(CH), (EH), AND (AH) ARE TOO DEMANDING

One might object that these modules of humility are too demanding as stated because they range over all people, and not just those with whom we have special relationships. It is difficult enough for human beings to exemplify humility in a consistent manner within special relationships, such as marriage, the parent-child relationship, and friendship. But these modules of humility are not restricted to these relationships; they extend to others as well. This underscores the demanding nature of this virtue and the objection at issue. Is the humble person to sacrifice all of her goods until the level of well-being of others is raised to an acceptable minimum, or hers is lowered to the point at which she has no more to give to satisfy the interests of others? Adopting a perspective with such universal scope is at best very difficult, and at worst nearly impossible.

In reply, it should first be pointed out that humility should not function in isolation from other virtues. In this context, as in many others, *phronesis*, or practical wisdom, will be relevant. For Aristotle, *phronesis* is "a state grasping the truth, involving reason, concerned with action about things that are good or bad for a human being."[29] Spiegel notes that this sort of "practical moral insight is essential in all moral conduct and decision-making."[30] In her account of *phronesis*, Linda Zagzebski helpfully characterizes this trait as a mediating virtue with respect to the intellectual and moral virtues. Such a virtue is needed because more than one virtue may be relevant to a particular set of circumstances. For example, both courage and humility may be relevant in a particular set of circumstances. Some features of the circumstances faced by the agent may relate to courage, while other features relate to humility. In such cases, virtues may appear to conflict with one another:

---

[28] There is much controversy about whether such partiality is morally permissible. For example, see Nagel (1991); and Feltham and Cottingham (2010).

[29] Aristotle (1999): 89/1140 b 5.

[30] Spiegel (2012): 56. Spiegel also discusses Zagzebski's account of this virtue.

> Suppose there are salient features of the situation that pertain both to courage and to humility. It might be that in *this* case humility is more important than courage. That is to say, it may be more important to act in a humble manner than to do what courage would have dictated in the absence of considerations of humility. Or it might be that humility is not so important in this case, although this is not to say that humility is *usually* less important than courage. The person who knows how to act (and feel) in these cases does not simply act from a combination of humility and courage and the knowledge of the proper function of these virtues. If we are to avoid an excessive and unworkable fragmentation of value, there needs to be some virtue that permits a person to sift through all the salient features of the situation—that is, all those features that are pertinent to *any* of the virtues—and to make a judgment that is not simply the judgment of a person qua courageous . . . or qua humble but is the judgment of a *virtuous* person. Therefore, the ability to mediate between and among the individual moral virtues must itself be a virtue. *Phronesis* is defined in part as the virtue that has this function.[31]

This should disavow us of the notion that the modules of humility are to be applied as a set of principles which can guide a fixed decisionmaking procedure yielding the correct course of action. It is the case that (CH) describes a cognitive perspective that the humble person possesses and (EH) a preference that she will have, and that she will possess (AH) which may yield the relevant humble acts as well. Nevertheless, there will be times when the humble person puts *her own interests* ahead of the interests of others, as prescribed by *phronesis*. Humility's modules are not virtue ethical analogues to Kant's categorical imperative or Mill's principle of utility, insofar as it will not be the case that the humble person must *always* put the interests of others before her own. Rather, she will rely on other virtues, including *phronesis*, in a manner that is sensitive to circumstances. This is important for both philosophical and practical reasons.

To see how this is so, consider the following case. How will humility dispose its possessor with respect to those who live in extreme poverty? The ideally humble person will believe (CH) and possess (EH) with respect to all of the world's poor, but on my account this does not entail that she must neglect her own interests as she seeks to come to their aid. Humility is consistent with a concern for one's own interests, while it also leads its possessor to be concerned for the interests of others. So the humble person may sacrifice some of her financial resources to come to the aid of some who suffer from grinding poverty, but she need not take the entire world on her shoulders. In a Christian context, it is important that she do her part, as a member of the body of Christ—that is, the universal church—while

---

[31] Zagzebski (1996): 222.

accepting that others have their roles to play as well. The Christian quest for social justice (and other goods) is a communal and not merely individual quest. This also entails that the humble person will allow others to make sacrifices on her behalf and for her good, as dictated by practical wisdom. An individual can be humble, believe (CH), and possess (EH), and yet allow others to put *her* interests ahead of *their* own. It is often pride that prevents one from accepting help or assistance from others, not humility, and all three of the modules under discussion, rightly understood, are consistent with this fact.

It is also important to point out that we can accept the morally demanding nature of these modules of humility while recognizing that in a Christian context, the individual can take herself to have many resources from which to draw as she seeks to adopt and exemplify humility. As a human made in the image of God, one grows in virtue in part by depending on the gracious and powerful assistance of God. In different ways, others can help in this as well. This idea is present within the context of the passage in Philippians. Prior to Philippians 2:3, which contains the admonition to consider others to be "more significant than yourselves," there are several background assumptions made by Paul which help ground the expectation that adopting such a perspective is in fact possible. The beginning of the passage assumes certain antecedent conditions prior to the transition to the moral exhortations it contains, including an experience of encouragement in Christ, the consolation of love, fellowship in the Holy Spirit, affection, and compassion. Paul is assuming that there is an ongoing moral and spiritual transformation within individuals that is in part a result of the power of God's Spirit in their lives, and is also a result which is partially dependent upon the existence of a certain form of community marked by affection, compassion, and unity. When these conditions obtain, (CH), (EH), and (AH) are less demanding than they would otherwise be, because we are not left on our own to cultivate and exemplify these modules of humility.

Finally, a qualification is needed. When the humble person's interests conflict with the interests of another, from her perspective, the burden of proof will rest on putting her interests ahead of the other person's as dictated by *phronesis*. In this way, Christian humility is demanding. In fact, the humble person may be called upon to sacrifice her very life, if the goods at stake merit such a sacrifice. The humble follower of Christ may be called upon to imitate her exemplar by making such a sacrifice. However, given the goods of union with Christ and exemplification of the character of Christ, if such a sacrifice is required it should not be seen as *too* demanding by those who seek to imitate him, even if it is the most difficult thing that one could be asked to do.[32]

---

[32] This chapter was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. I would like to thank Christian Miller, Angela

References

Adams, Robert. 2006. *A Theory of Virtue*. New York: Oxford University Press.

Aristotle. 1999. *Nicomachean Ethics*. 2nd ed. Translated by Terence Irwin. Indianapolis, IN: Hackett.

Audi, Robert. 2011. *Epistemology: A Contemporary Introduction*. 3rd ed. New York: Routledge.

Chisholm, Roderick. 1977. *Theory of Knowledge*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Copan, Paul. 2006. "Divine Narcissism? A Further Defense of God's Humility." *Philosophia Christi* 8: 313–325.

DeWeese, Garrett. 2007. "One Person, Two Natures: Two Metaphysical Models of the Incarnation." *Jesus in Trinitarian Perspective*, edited by Klaus Issler and Fred Sanders, 114–153. Nashville, TN: B&H Academic.

Dickson, John, and Brian Rosner. 2004. "Humility as a Social Virtue in the Hebrew Bible?" *Vetus Testamentum* 54: 459–479.

Feltham, Brad, and John Cottingham, eds. 2010. *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*. New York: Oxford University Press.

Fullam, Lisa. 2009. *The Virtue of Humility: A Thomistic Apologetic*. Lewiston, NY: Edwin Mellen

Hurtado, Larry W. 2005. *How on Earth Did Jesus Become a God?* Grand Rapids, MI: Eerdmans, 83–107.

Lewis, C. S. 1952. *Mere Christianity*. New York: Macmillan.

Nagel, Thomas. 1991. *Equality and Partiality*. New York: Oxford University Press.

Nussbaum, Martha. 2001. *Upheavals of Thought*. New York: Cambridge University Press.

Pardue, Stephen T. 2013. *The Mind of Christ: Humility and the Intellect in Early Christian Theology*. London: T&T Clark.

Pettit, Philip. 2006. "Preference, Deliberation and Satisfaction." *Preferences and Well-Being*, edited by Serena Olsaretti, 131–153. Cambridge: Cambridge University Press.

Piller, Christian. 2006. Content-Related and Attitude-Related Reasons for Preferences. In *Royal Institute of Philosophy Supplement* 59: 155–181.

Richards, Norvin. 1992. *Humility*. Philadelphia, PA: Temple University Press.

Roberts, Robert C., and W. Jay Wood. 2007. *Intellectual Virtues*. New York: Oxford University Press.

Spiegel, James. 2012. "Wisdom." *Being Good: Christian Virtues for Everyday Life*, edited by Michael W. Austin and R. Douglas Geivett, 53–71. Grand Rapids, MI: Eerdmans.

———. 2003. "The Moral Irony of Humility." *Logos: A Journal of Catholic Thought and Culture* 6: 131–150.

Strimple, Robert. 1979. "Philippians 2:5-11 in Recent Studies: Some Exegetical Conclusions." *Westminster Theological Journal* 41: 247–268.

Taliaferro, Charles. 1989. "The Vanity of God." *Faith and Philosophy* 6: 140–154.

Knobel, Cristian Mihut, Nancy Snow, James Spiegel, Ian Church, and Trent Dougherty for their comments on previous versions of this chapter. I presented versions of it at the Character Project Conference, Wake Forest University, June 28, 2013, and at the Talbot Philosophical Society, Biola University, February 12, 2013. I received helpful comments from those in attendance at both presentations, for which I am grateful.

Verbrugge, Verlyn. 2000. *New International Dictionary of New Testament Theology*. Abridged ed. Grand Rapids, MI: Zondervan.

Whittaker, John. 1992. "'Agape' and Self-love." *The Love Commandments: Essays in Christian Ethics and Moral Philosophy*, edited by Edmund Santurri and William Werpehowski, 221–239. Washington, DC: Georgetown University Press.

Wielenberg, Erik J. 2005. *Value and Virtue in a Godless Universe*. New York: Cambridge University Press.

Zagzebski, Linda. 1996. *Virtues of the Mind*. New York: Cambridge University Press.

16 }

A Different Kind of Wisdom

Angela Knobel

Recent scholarship has done much to undermine any easy identification of the respective virtue theories of Aquinas and Aristotle.[1] In what follows I will examine one difference in particular—namely, how Aquinas's Christian commitments lead him to offer a rather un-Aristotelian picture of the activity of practical reason in the virtuous Christian. Specifically, I will argue that while Aristotle's virtuous pagan is marked by *self-sufficiency* in moral reasoning, Aquinas's virtuous Christian is marked by his *dependency* in the same.[2] Because it is Aquinas's commitment to fundamental Christian notions that generates this difference, his distinctive account of Christian practical reason will be of interest even to those Christians who do not share his other philosophical commitments.[3]

In what follows, I will first describe the sense in which Aristotelian and neo-Aristotelian[4] virtue ethics presupposes what I will call the *self-sufficiency* of the virtuous agent's practical reason. I will argue that this presupposition departs in important ways from the virtuous Christian's own self-understanding. I will

[1] Philippa Foot's remark, in her 1979 essay "Virtues and Vices," that Aquinas followed Aristotle "heroically" would not have seemed at all controversial to her readers. Indeed, Fredrick Copelston makes a similar, if more nuanced, claim in his summary of Aquinas's moral theory, saying that Aquinas "utilises the Aristotelian ethic but fits it into a Christian setting." Remarks like this are in stark contrast to accounts of contemporary scholars, such as Jean Porter, who argues that Aquinas "subverts" Aristotle's ethics. For statements of the traditional view, see Foot 1997: 163–177; and Copelston 1962: 123. For examples of recent arguments to the contrary, see Porter 1992: 19–41; and Stump 2011: 29–43.

[2] A more precise definition of "self-sufficient" will be offered in what follows.

[3] I say other "Christians" because—as I will show—Aquinas's views about the dependency of Christian practical reason stem directly from his recognition of the role that the Holy Spirit plays in the Christian moral life. Insofar as a religion does not recognize a similar role for the Holy Spirit or even the Holy Spirit at all, Aquinas's account would be less impactful.

[4] Hereafter, "Aristotelian" will be taken to include both traditional Aristotelian and neo-Aristotelian accounts of the virtues.

then contrast this with Aquinas's understanding of the operation of Christian practical reason. While Aquinas may describe pagan practical reason in much the same way that Aristotle does, I will argue that he understands Christian practical reason to be fundamentally *dependent* and, moreover, that the virtuous agent's recognition of this dependency is a mark of moral maturity.

1. Posing the Problem

In this section I will describe one way in which Aristotelian accounts of virtue imply a picture of the moral life which departs in important ways from the virtuous Christian's own self understanding. Specifically, I want to focus on something implied by the Aristotelian account of *phronesis*.[5] I will argue that—according to the Aristotelian account—the various features of practical wisdom[6] imply that the virtuous agent's self-sufficiency as a moral reasoner increases apace with his virtue. That is to say, under an Aristotelian account of virtue, the more virtuous an agent is, the less he needs to seek the assistance of others in determining the appropriate course of action in a given situation. This notion, I will argue, runs contrary to the self-understanding of the virtuous Christian.

In an Aristotelian virtue ethic, *phronesis* (practical wisdom) is not merely a virtue in its own right but an integral component of all the moral virtues as well. Not least because virtue theory eschews the notion that right action consists in the following of predetermined rules, the virtue theorist must offer some account of how a virtuous agent locates the appropriate action in a given situation: that is, even if the virtuous agent *wants* to do the courageous thing, how does one determine what the courageous thing in a given situation *is*? On the standard Aristotelian account, the virtuous agent is able to determine the appropriate (that is to say, virtuous) response to a given situation because to possess virtue is also to possess practical wisdom.[7] While it is through moral virtue that one desires the appropriate ends, it is through practical wisdom that one is able to determine how those ends should be pursued in a particular situation. Julia Annas, for instance, describes practical wisdom as "a disposition to decide and act that involves the kind of understanding *only you* can achieve in your own case."[8]

---

[5] Since my thesis is that Aquinas departs from Aristotle in important ways precisely where practical wisdom is concerned, it is appropriate to demonstrate that Aquinas's account of practical wisdom does not have this implication. I will address this point in what follows.

[6] By "practical wisdom," I mean the same thing that Aquinas and other scholars mean by "prudence." For the sake of simplicity, this chapter refers only to "practical wisdom."

[7] For an excellent discussion of the importance of practical wisdom for virtuous action, see Russell 2008: 337. Russell offers a more extensive account of practical wisdom in his *Practical Intelligence and the Virtues*. See Russell 2009. See also Hursthouse 1999: 56–59; and Aristotle 1999: book 6.

[8] Annas 2005: 518. Although Annas would describe herself as a Stoic virtue ethicist more than an Aristotelian virtue ethicist, the difference is not relevant to the present point.

The individual who possesses *phronesis* grasps the morally relevant features of a given situation, and he does so on his own. The qualification "on his own" admittedly requires clarifying. What I have in mind by "on his own" is what is implied by Annas's assertion that practical wisdom "involves the kind of understanding *only you* can achieve." It is not that the possession of *phronesis* implies that one never need rely on anyone or anything. We all rely on other things and people for many things, in many areas of life, and things would surely be no different in the moral sphere. The claim is not that the individual with *phronesis* does not seek information from other people or rely on their assistance, but that he is the one who ultimately *recognizes* the morally relevant features of the action: someone else does not "feed" him the appropriate assessment of the situation at hand.

For instance, while one might wish to be kind or generous, it is often unclear what the kind or generous thing to do is. Do I really help the homeless person by giving him my spare change, or does such an action actually do more harm than good? Is it a kindness to let my child discontinue the piano lessons she loathes, or does true kindness consist in insisting that she continue them? Both those who possess *phronesis* and those who lack it might well seek more information: both might find out what the child hates about the lessons, or discuss the child's aptitude with the instructor, say; but when all is said and done, the person with *phronesis* is distinguished by his ability to grasp the appropriate course of action to take in the particular situation. To possess *phronesis* is in part to possess the ability to accurately evaluate situations like these. The prudent mother has the situation-specific insight that enables her to make good choices about childrearing, and her insight resists formulation into any general rule: were the child or the teacher or some other feature of the situation different, she might well arrive at a different decision.

Especially because the insight that practical wisdom presupposes cannot be formulated into general rules, practical wisdom also seems to presuppose experience. A mother who has raised six children often does seem to understand more about when it is (or isn't) appropriate to insist on piano lessons than a mother struggling to raise her first child. In the normal course of things, the only way to cultivate such insight is through experience: one must gradually learn what does and does not work, or at least seek the advice of someone who already has the requisite insight. Thus, Aristotle says that while one finds young people who are gifted at mathematics and music, one does not find young people who possess *phronesis*.[9] Contemporary virtue ethicists echo this point.[10] The important point in all of this is not merely that Aristotelian practical wisdom presupposes experience, but that *phronesis* is a skill that has to be

[9] Aristotle 1999: 6.1142a15.

[10] It is this observation, for instance, that leads Linda Zagzebski to argue that it would be logically impossible for an individual to be "given" a virtue all at once. See Zagzebski 1996: 120–122.

cultivated. The insight practical wisdom requires is acquired gradually, over a considerable period of time.

This description indicates another feature of *phronesis*, namely, that it is an ability that can be possessed in degrees. Even if Aristotle himself did not make such a claim, it is only reasonable that this would be the case. Especially because *phronesis* seems to presuppose experience, it stands to reason that individuals would gradually increase their ability to exercise practical reason: a mother will understand more about parenting with her second child than with her first, and so on.

To possess the practical reason that makes virtuous action possible, then, is to possess a kind of insight that (a) is unique to one's own self, situation, and circumstances; (b) is cultivated over a substantial period of time; and (c) exists in degrees. None of these claims is particularly controversial for the Aristotelian. They do, however, imply something about what it means to possess *phronesis*, and it is this that I want to focus on.

The foregoing description of *phronesis* implies that the more *phronesis* an individual possesses, the more *self-sufficient* that agent's moral reasoning will be. Since the term *self-sufficient* is admittedly vague: it is important to clarify my meaning. By *self-sufficient* I mean four things. First, the individual who possesses Aristotelian *phronesis* possesses an insight that makes him uniquely suited to recognize the morally relevant features of his own experiences. In our previous example, the practically wise mother was able to recognize features of the situation a less wise mother would not: she recognizes aspects of her child and of the particular situation that others would not have. More importantly, however, the practically wise mother's insights are not widely available: only she is capable of seeing what she sees. Second, as is indicated by the emphasis on the prerequisite of experience, *phronesis* is a skill that must be cultivated by—and only by—the agent himself. One might well seek and even require assistance in cultivating the skill, but the skill is nonetheless something that the agent himself must cultivate over time.[11] To develop *phronesis* is to become adept at recognizing the morally relevant features of a situation, and this is something that agent himself must develop. Third, one's ability to recognize the morally relevant features of situations corresponds to the degree of *phronesis* one possesses. This, indeed, seems to be the insight conveyed by the consistent claim that one cannot possess *phronesis* unless one has experience. Together, these criteria imply a fourth—namely, that an individual's confidence in his own moral assessments of situations should grow apace with his virtue. Under the Aristotelian account, a virtuous individual should possess a confidence in his ability

[11] One might well draw analogous comparisons with the skills required to play various sports. The guidance and instruction of a coach certainly helps a great deal, but no amount of instruction and guidance can *create* the skill in the athlete. The instructor can only guide the player's own efforts.

to recognize the morally relevant features of a situation that corresponds to the degree of virtue he possesses.

While this account of *phronesis* should seem relatively uncontroversial to Aristotelians, it also runs contrary to lived experience in a crucial way: it does not adequately describe the self-understanding of the virtuous person of faith. It is characteristic of committed members of all monotheistic faiths to ask God to provide them with the insight that the *phronesis* described above is supposed to yield and—I submit—the more a deeply religious person seeks to live a morally upright life, the more insight he asks God to give him. This is particularly evident in the Christian tradition. Proverbs 3:5–6, for instance, enjoins trust in God over one's own understanding: "Trust in the Lord with all thine heart; and lean not unto thy own understanding."[12] When he sends his disciples out into the world, Christ likewise tells them not to worry about what to say when they are arrested and imprisoned, telling them that the Holy Spirit will provide them with the words they need.[13] Numerous biblical figures likewise receive (sometimes unsolicited) signs—either from God or from other people—signs that often conflict with their own assessment of the situation.[14] These biblical examples are entirely consistent with the committed Christian's own self-understanding.

At least in the Christian tradition, then, it is considered entirely fitting for a virtuous person of faith to ask for God to provide him with insight rather than to rely on his own assessment of the situation. What are we to say about this? One possibility is to simply deny that deeply religious people do in fact possess any high degree of practical wisdom: their reliance on divine assistance is indicative of poorly or insufficiently developed practical reason. Indeed, insofar as virtue is necessarily accompanied by *phronesis*, and insofar as *phronesis* is in turn accompanied by the types of self-sufficiency just described, such individuals are simply not virtuous. Another possibility would be to deny any necessary correspondence between proficient practical reason and virtue.[15] Or, we could say that such individuals are simply misguided: they actually do possess *phronesis*, but don't realize it. In such cases, what one imagines to be God-given insight would simply be the result of one's own careful deliberation.

In what follows, I want to propose a different hypothesis, which I think Aquinas's account of Christian virtue exemplifies. I will argue that in the virtuous person of faith, virtue goes hand in hand with *phronesis*, but that the *phronesis* relevant to the life of faith is not constituted by the kind of self-sufficiency I have just described. To the contrary, the more virtuous a person of faith is, the more that individual both recognizes the need for and relies on divine assistance in the exercise of practical reason. Although I think this claim is true for members of

[12] *Holy Bible* 1963: Proverbs 3:5–6.
[13] Ibid.: Matthew 10:19–20; Mark 13:11.
[14] Examples include Jonah, Zachariah, Abraham, and Mary, among others.
[15] See, for instance, Slote 1997.

any number of religious traditions, I will focus on the Christian religion in particular because it is this tradition to which Aquinas's remarks are most pertinent.

2. Two Ends

When I posed the problem earlier, I deliberately neglected a crucial difference between a standard Aristotelian (i.e., eudaimonist) virtue ethic and the specifically Christian virtue ethic envisioned by Aquinas. I claimed that the virtuous Christian has rather different views about the autonomy of practical reason. But I glossed over a more central, more important difference. Virtues are so called because they not only enable but are partially constitutive of flourishing. If the Christian has a decidedly different view of flourishing, then it is likely that he will also have a different conception of the virtues that make flourishing possible, and it is even possible that he will have a different view of the role of practical reasoning in the activity of those virtues.

All of this is particularly relevant to Aquinas, because he recognized two distinctively different types of flourishing: an Aristotelian, this-worldly kind of flourishing, on the one hand, and a type of flourishing only possible for the Christian, on the other.[16] That he privileged the latter should not be taken to diminish the importance of the former. In many ways, Aquinas's understanding of the kind of flourishing possible outside the Christian life both sets the stage for and helps one make sense of his view of Christian flourishing. This is particularly true of his view of the different roles of practical reason in each type of flourishing. In what follows, I will contrast Aquinas's views of natural and supernatural flourishing. I will focus specifically on the contrast between (1) the activity of practical reason at the natural and supernatural levels, and (2) natural and supernatural moral growth.

2.1 NATURAL VERSUS SUPERNATURAL FLOURISHING

Even though Aquinas believed that man's true fulfillment is a supernatural one, he acknowledged that a lesser kind of flourishing could be pursued apart from revelation and grace.[17] When Aquinas describes this lesser kind of flourishing,

[16] As more than one scholar has noted, Christian scholars prior to the eleventh century almost unanimously rejected the notion that a pagan could possess virtue. Around the eleventh century, for a variety of reasons, this began to change. By Aquinas's time, many scholars agreed that non-Christians could achieve a kind of virtue, and they agreed that such virtues were genuine virtues; they simply weren't the highest form of virtue that could be had. (For a lengthier discussion of this, see Bejczy 2005.) Aquinas's description of these "pagan virtues" is very close to Aristotle's and, in some respects, even fuller: Aquinas fills in details that Aristotle omits. In what follows I'm going to say something about Aquinas's view of natural virtue—that is, about the kind of virtue he thinks even a non-Christian can pursue.

[17] Aquinas 1964: I–II q.5 a.3.

he describes the relationship between virtue and practical reason in a way that sounds very Aristotelian, so much so that some scholars claim that Aquinas simply adds additional details to Aristotle's already existing framework.[18] For the sake of clarity, I will refer to the type of flourishing that Aquinas believes to be possible apart from revelation and grace "natural flourishing," and the virtues and practical reason that are constitutive of it "natural virtue" and "natural practical reason."

Since Aquinas held that man attains natural flourishing by exhibiting the excellences proportionate to his nature, and since he, like Aristotle, held that one exhibits the excellences proportionate to one's nature when one exhibits right reason in action, Aquinas agreed with Aristotle that one pursues natural flourishing by cultivating developed capacities to act in accord with right reason—that is, by cultivating natural virtues.[19] Although Aquinas also agreed with Aristotle that natural virtues go hand in hand with natural practical reason, he also had a rather more developed account of the process through which these traits are cultivated.

Aristotle's account often seems to make natural virtue highly dependent on chance: someone who has the misfortune of a bad upbringing or unfortunate birth might not be able to be virtuous. Aquinas, though, holds that all men naturally possess a basic moral sense of direction.[20] That is to say, he held that all men, simply in virtue of their membership in the species, possess both the "beginnings" of natural virtue and the resources to cultivate the natural virtues in themselves.[21] These beginnings, for Aquinas, are man's natural knowledge of the first moral principles and his natural desire for the good of reason.[22] By "first moral principles" Aquinas means the practical equivalents of the first principles of speculative reason. For instance, Aquinas holds that speculative reasoning takes the principle of non-contradiction for granted, and that this assumed principle is in fact a precondition of any further speculative reasoning. In a similar vein, he holds that "the good must be done and evil avoided" is assumed in any practical reasoning and, again, is a precondition of it. If man did not operate from the principle that good must be done and evil avoided, and if he did not, moreover, naturally desire what reason presents to him as good, practical reasoning would not be possible. Just as all theoretical knowledge derives from certain naturally known first principles, so, Aquinas

[18] Foot 1997: 163–177. More recently, Candice Vogler has made a similar albeit much more nuanced claim. See Vogler 2012.

[19] For Aristotle, of course, this was a necessary but not sufficient requirement. See Aristotle 1999: book I.

[20] Since this natural direction is present as part of man's very constitution, it would be potentially present even in infants, though it could of course not be exercised until the individual began to reason. Aquinas does hold, however, that it is present even in the earliest exercise of reason.

[21] Aquinas 2005: a.8.

[22] Ibid.; see also Aquinas 1964: I–II q.63 a.2.

believed, practical reason takes its departure from naturally known moral principles, first among which is "do good and avoid evil."[23]

Since all men naturally possess foundational moral knowledge and a desire for the good of reason, and since all men also possess reason, all men can in principle exercise natural practical reason correctly, but only the virtuous can do so consistently. Just as anyone who knows Euclid's five postulates can, in principle, reason to correct conclusions about geometry, so all men can, in principle, reason to the correct action in a given situation and act accordingly. Moral beginners, just like those unschooled in geometry, will often need help in their moral reasoning (who of us, after all, learned geometry by deriving it from the five postulates?). All the same, Aquinas thought that even when men receive assistance in acts of natural virtue, the one who provides assistance must minister to their reason: she helps their reason follow the course it would take if it were fully developed.[24]

When one reasons (perhaps with assistance) from one's natural moral knowledge to a correct conclusion and acts accordingly, one not only performs an act of natural practical reason; one also performs an act of natural virtue. All of this can occur without actually possessing the virtue in question; indeed, it is via such actions that the natural virtues come to be. When one manages to perform such actions repeatedly, one simultaneously cultivates both natural moral virtue and natural practical reason. And, of course, the more proficient one's natural practical reason becomes, the less assistance one needs. The diagram in figure 16.1 helps to illustrate this:

Aquinas does not, of course, think that the pursuit of natural virtue is easy. Thanks to the Fall, human beings experience many obstacles, even in their pursuit of natural flourishing.[25] Nonetheless, it is clear that Aquinas thought that man could pursue natural flourishing through his own natural resources, and it is likewise clear that on this picture of natural flourishing, the degree to which one is proficient in practical reason corresponds to the degree of virtue one possesses. This picture likewise gives us every reason to think that proficient natural practical reason is autonomous: one possesses the natural virtues to the degree to which one can reason—"on one's own"—to the correct course of action in a given situation. By "on one's own" here I mean that on the Thomistic account increased proficiency, as a moral reasoner implies increased proficiency at reasoning from the first moral principles to the conclusions that derive from them. The less proficient a moral reasoner is, the more assistance he will require in order to see what a situation morally requires, just as less proficient mathematicians need a teacher to help them reason through a

[23] Ibid.: "[T]hese are known naturally, and are the principles on which subsequent knowledge . . . depends, whether it is practical or theoretical."

[24] Aquinas 1953: q.11.

[25] Aquinas 1964: I–II q.82.

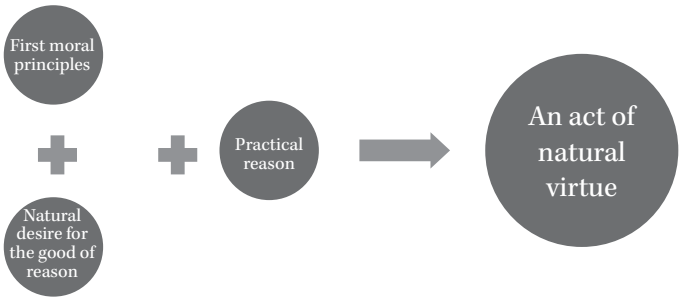

FIGURE 16.1 An Act of Natural Virtue.

calculation. There is no need to read Aquinas as claiming that the morally proficient agent would not ever, say, look up a word in a dictionary. The point is merely that when an agent is a proficient moral reasoner, then once he grasps all the morally relevant features of the situation, he readily grasps what that situation requires. When an agent lacks that proficiency, he might have similar knowledge of all the morally relevant features and yet still be unable to grasp what the situation requires of him.

2.1.1 Supernatural Flourishing

In the preceding, I showed that Aquinas believed that all men can pursue natural flourishing because all men are born with both an initial orientation to it and the resources necessary for its pursuit. But for Aquinas, of course, true fulfillment is rather different. True fulfillment consists in participation in the divine life, something entirely outside the human being's grasp. No matter how adept one becomes at cultivating one's natural capacities, one can never achieve union with God. The only truly important kind of flourishing, then, requires an elevation of human nature itself. This elevation of human nature is achieved through the gift of God's grace, but Aquinas has a unique understanding of how it is that grace affects the human being who receives it.

Aquinas calls grace a habit, but it is vital to understand that he does not consider it to be just one habit among many. Under the traditional Aristotelian account, a *habit* is a disposition that either helps man achieve his end (in which case, it is a virtue) or hinders that pursuit (in which case, it is a vice). Under the traditional Aristotelian account, man's end itself is stable. But on Aquinas's account, the effect of the gift of the habit of grace is the elevation of man's nature, so that man is now capable of directing his actions to an end—participation in the divine life—to which he was not previously capable of directing his actions. For this reason, Aquinas claims that grace is a habit that inheres in the very essence of the soul, instead of in one of its powers. While a good habit inhering in one of the soul's powers enables that power to produce acts ordered to man's end, a habit of the essence of the soul (grace) causes man's end itself to be elevated. Instead of a soul ordered to

natural Aristotelian flourishing, the soul perfected by the habit of grace is now ordered to participation in the divine life. Explaining this point, Aquinas says, "[I]f the nature in question is a superior nature in which a human being may come to share, in the way in which St. Peter speaks of us being made sharers in the divine nature," then the soul itself needs a habit—namely, grace.[26] Aquinas believes that grace transforms human nature, enabling man to participate in the divine life. Even if grace is a habit, then, it is not one habit among others, but a habit that makes the entire Christian life possible.

The elevation that the habit of grace effects in the soul goes, so to speak, all the way down, even down to our basic orienting moral principles. Because of this, Aquinas hypothesizes that God infuses a number of other habits along with grace. These latter habits perfect the powers of the soul, thus enabling those powers to produce acts ordered to union with God. We saw earlier that—for Aquinas—all men have a basic ordering to the good of reason insofar as all men possess basic habitual moral knowledge and insofar as all men naturally desire the good of reason. At the supernatural level, however, this basic ordering is insufficient. Man needs a different basic moral direction in order to be properly ordered to supernatural beatitude. It is for this reason, says Aquinas, that when man receives the gift of habitual grace from God he also receives the infused theological virtues of faith, hope, and love.[27] These virtues play the same role at the supernatural level that man's natural moral knowledge plays at the natural level: they give man his first, initial order to participation in the divine life.[28] In addition to these "foundational" virtues, Aquinas also holds that God bestows infused versions of all the moral virtues (i.e., infused courage, infused temperance, and so on) along with the habit of grace. We will return to these momentarily.

But what about practical reason at the supernatural level? We saw that man pursues natural flourishing by using his reason to move from his basic natural moral knowledge to correct conclusions about how to act. When he does so repeatedly, he cultivates the virtues, and the more virtuous he becomes, the more easily he reasons to the appropriate action. When he is less virtuous, he needs a teacher: someone to help his reason follow the appropriate path. As he becomes more virtuous, his reason is able to move along the appropriate paths without assistance. If we were to follow the analogy through, we might be inclined to say that at the supernatural level, reason does the same thing but simply begins from different starting points. Instead of the knowledge of natural law, reason starts with faith, hope, and love and reasons from those starting points to a correct conclusion about what is to be done. Although

[26] Aquinas 1964: I–II q.50 a.2.
[27] Aquinas 2005: a.10; Aquinas 1964: I–II q.63 a.3.
[28] Aquinas 2005: a.10; Aquinas 1964: I–II q.63 a.3.

scholars have historically maintained[29] that this is Aquinas's view, I believe that Aquinas's actual position is rather different.

Although Aquinas believes that faith, hope, and love give man a supernatural sense of moral direction analogous to the natural moral direction he receives from his natural knowledge of first moral principles and his natural desire for the good of reason, he does not believe that the former guide man as reliably as the latter. Although the divinely given virtues of faith, hope, and love are far greater than man's knowledge of the natural law, he also possesses the former much less perfectly than the latter.[30] Because of this, man's ability to reason from these starting points to those acts that befit his participation in the divine life is impaired. Since Aquinas is so blunt about the insufficiency of reason at the supernatural level, it is worth quoting him at length:

> [M]an is able to act by the judgment of reason in regard to matters that are subject to human reason, viz. to perform acts ordained to man's connatural end . . . But the case of the ultimate and supernatural end is different. Towards it man is moved by reason in so far as reason is formed by the theological virtues, which form it only after a fashion and imperfectly.[31]

Aquinas offers two dramatic examples to illustrate the insufficiency of human reason in the Christian moral life. Man needs God's assistance in order to perform acts befitting his participation in the divine life, Aquinas says, in the same way as the moon cannot give off light without the sun, and as the medical student cannot practice his art without the guidance of a doctor.[32] The examples are striking for the extent of the dependence they imply.

Aquinas believes that it is the Holy Spirit that provides the Christian with the guidance he needs. With grace, men receive the gifts of the Holy Spirit (wisdom, understanding, counsel, knowledge, fortitude, piety, and fear[33]), which make men receptive to the "motion" or "inspiration" of the Holy Spirit.[34] Perhaps most importantly, when Aquinas speaks of man's need for the guidance of the Holy Spirit, he uses the word *always*.[35] It is not that man sometimes or even often needs divine assistance to navigate the Christian moral life; it is that he *always* needs that assistance. Even an act of courage or temperance

---

[29] Reginald Garrigou-Lagrange, for instance, offers such an account. He acknowledges that divine inspiration can sometimes supersede this but does not think such inspiration is typical. Although Aquinas's early *Scriptum Super Sententiis* does indicate such a position, I think it is clear that Aquinas had changed his position by the time of his *Summa Theologiae*. See Garrigou-Lagrange 1939.

[30] Aquinas 1964: I–II q.68 a.2.

[31] Ibid.

[32] Ibid.

[33] This list is not, of course, unique to Aquinas. The fathers of the church uniformly acknowledged the same list, though Aquinas's explanation of the role of the gifts is unique.

[34] Aquinas 1964: I–II q.68 a.2

[35] Ibid.

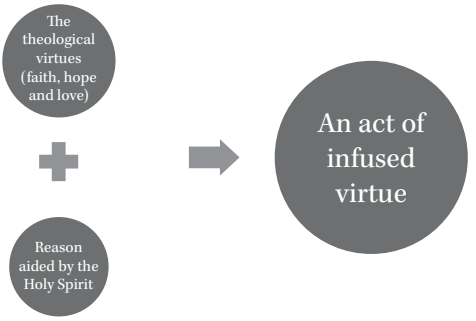

FIGURE 16.2 An Act of Supernatural Virtue.

will no longer be the same as it would be in the non-Christian. In Aquinas's view, the Christian's courageous act will be an act of infused courage. Without going into all the details here, we can say that the correct reasoning necessary to the act of infused courage is only possible through the assistance of the Holy Spirit. It is only thanks to the Holy Spirit's inspiration and guidance that man is able to successfully navigate the Christian moral life. As opposed to the purely natural level, then, virtuous supernatural action would look like figure 16.2:

It is crucial to note that although Aquinas believes that man cannot navigate the Christian moral life without the guidance of the Holy Spirit, the Holy Spirit's guidance does not in anyway negate or eliminate the importance of human reason, and it does not diminish man's free will. The Holy Spirit assists man by helping him to correctly appraise the situation at hand. Aquinas makes this claim explicitly,[36] but the interaction of human reason and the Holy Spirit are best understood through his example of how the gift of counsel assists the activity of practical wisdom.

Aquinas believed that man receives assistance from the Holy Spirit through the "gifts," and he believed that each gift enables man to receive a different form of assistance. The gifts are not, in Aquinas's view, virtues. Rather, the gifts are bestowed because even those who possess grace and the infused moral and theological virtues *still* need divine assistance, assistance which the gifts of the Holy Spirit enable them to receive. The gift of counsel, for instance, enables man to receive guidance from the Holy Spirit in his practical reasoning. Through the gift of counsel, the Holy Spirit helps guide man's deliberation.[37] The gift of counsel is necessary, says Aquinas, because man cannot grasp all the particulars of a given situation and, quoting scripture, because "the thoughts of mortal men are fearful and our counsels uncertain."[38] The inspiration the Holy Spirit provides through the gift of counsel does *not*, importantly, force

[36] Aquinas 1964: II–II q.52 a.1 ad.3
[37] Aquinas 1964: II–II q.52 a.1
[38] Ibid.

man's hand: it does not compel him to make one decision over another. But it does enable him to recognize features of the situation he otherwise would not.

The assistance the Holy Spirit provides is not, importantly, limited to propositional knowledge. Just as in an ethic of flourishing, the moral virtues and practical reason are interdependent, so, Aquinas believes, Christian practical reason is interdependent with Christian moral virtues (which are not acquired but infused by God). And even the Christian's virtuous responses need the assistance of the Holy Spirit. The Holy Spirit can assist man by providing him with confidence, or by causing him to feel pity in the appropriate situation, and so forth.[39] Although I do not have the space to enter into details here, it is clear that Aquinas considers the guidance of the Holy Spirit to be a vital component of not only Christian practical reason but the Christian moral life as a whole.

One might think that the Christian only needs the guidance of the Holy Spirit temporarily: perhaps as the Christian continues to pursue supernatural beatitude, he will need the Holy Spirit less. If this were the case, the contrast between proficient practical reason at the natural and supernatural levels would be considerably diminished. This, though, is precisely not Aquinas's view. To the contrary, the more one grows in the Christian moral life, the more attuned and receptive to the inspiration of the Holy Spirit one becomes. In those in whom grace is only weakly present or whose disordered passions pose an obstacle, the guidance of the Holy Spirit is felt less. Growth in the Christian life, for Aquinas, means becoming more and more attentive to and receptive of the working of the Holy Spirit in oneself.

Although I will not dwell on this point here, Aquinas also believes that the virtuous Christian seeks and receives assistance from others. One might assume that even if he continues to need *God's* guidance, the morally virtuous Christian will be less in need of the assistance of others; his insufficiency is limited to divine things. Although this possibility certainly warrants examination, I do not think this is Aquinas's view. Indeed, Aquinas emphasizes that the virtuous Christian recognizes God's gifts in others, and he emphasizes the role of virtues like obedience and humility in the Christian life.[40]

3. Aquinas and Contemporary Christian Virtue Ethics

Aquinas's account of Christian practical reason is consistent with the accounts of Christian spiritual writers, who uniformly emphasize man's dependence on God. In his *Tuning in to Grace*, for instance, André Louf argues that Christian

[39] It is important to distinguish a "gift" of the Holy Spirit from what it is that the Holy Spirit actually provides. In Aquinas's view, the gift (counsel, wisdom, etc.) is merely a disposition in man that makes him *amenable* to the Holy Spirit's inspiration. The disposition that makes one receptive to the Holy Spirit's inspiration is not, of course, to be identified with the inspirations themselves.

[40] Aquinas 1964: I–II q.104 and q.161.

virtue "has as much to do with weakness as grace."[41] Acknowledgment of one's own weakness is necessary, Louf argues, because blindness to one's own insufficiency is the fundamental barrier to God. The life of grace is made possible by the recognition that we are weak and in need of help. The Curé de Torcy in Georges Bernanos's *Diary of a Country Priest* makes the same point, more succinctly, when he tells the country priest that "prudence is the final imprudence when by slow degrees it prepares the mind to do without God."[42] Louf is speaking of Christian virtue more generally, while Bernanos's quote has to do with practical wisdom in particular, but the intent of both is clear: to rely on and be confident in one's own strength, even or perhaps especially when it comes to practical wisdom—is to shut out God. In this chapter I have argued that Aquinas, with his emphasis on man's continual need for divine guidance, accommodates these insights nicely.

If Aquinas's account of Christian virtue is attentive to man's need for divine guidance, and if he rightly understands that the direction of the Holy Spirit is an integral part of Christian practical wisdom, then it would seem that his account has much to contribute to contemporary accounts of Christian practical wisdom. But incorporating Aquinas's insights into contemporary accounts also presents challenges of its own. I conclude by mentioning one of these challenges.

Aquinas's account of Christian practical wisdom contains a feature that is at the same time theologically intuitive and philosophically problematic: under Aquinas's account the virtuous Christian's practical wisdom itself is in an important sense not his own. The possessor of Aristotelian *phronesis* is characterized by his ability to recognize the morally relevant aspects of a situation, and the more accurate his insight is, the more confident he is in his assessment of the situation, the more practical wisdom we attribute to him. The Christian, on the other hand, consciously and continuously seeks God's guidance. On the account I have described, the virtuous Christian is always ready to doubt his own assessment of a situation and considers himself to be continually in need of God's guidance. And this might well seem problematic, especially if we also wish to claim that the Christian both has practical wisdom and that his practical wisdom is a *virtue*. Two objections are particularly relevant.

Both objections stem from the possibility that what this account describes as Christian "practical wisdom" is not, at the end of the day, a virtue at all. First, virtues are good habits, and as such, seem to be the sorts of things that by their very nature must be acquired through an individual's own repeated good actions. What I have called Christian practical wisdom, though, is by its very definition *not* something that one can acquire through one's own efforts. So why call it a

[41] Louf 1992: 48.
[42] Bernanos 2002: 91.

virtue at all? Second, when we say that an individual possesses a virtue, we ascribe something praiseworthy to him. But even assuming that it *is* possible (contra the first objection) to possess a virtue that one did not acquire, how could the agent really deserve to be praised for possessing it? Surely, we should not praise an individual for possessing what he did not earn? Neither of these objections can be answered fully here, so I will close with some brief thoughts about how each objection might be addressed.

More than one contemporary virtue ethicist has claimed that virtue is by its very nature something that an individual must acquire through his own repeated good actions.[43] Such an objection, if true, would make an account like Aquinas's impossible. For the Christian practical wisdom described above is a gift through and through. It is through God's gift that man receives the grace that elevates his nature, through God's gift that man receives the capacities (i.e., the theological and infused moral virtues) that enable him to act in accord with his elevated nature, and finally, through God's gift (i.e., the gifts of the Holy Spirit) that man is made capable of receiving the assistance of the Holy Spirit in his practical reasoning. Christian virtue in general and practical wisdom in particular is, on Aquinas's account, a gift. If the contemporary claim that virtue must be acquired is correct, then Aquinas's account cannot even get off the ground.

Although the above objection cannot be fully answered here, it is worth noting that the claim is nowhere nearly as obvious as it might initially seem. Even for Aristotle, a virtue is first and foremost an excellence: it enables a thing to act in accord with its nature.[44] Although Aquinas certainly believes that human nature must be elevated by grace, the Christian virtues he describes do precisely this: they enable man to act in accord with a nature that has been perfected by grace. Under the traditional Aristotelian account, it is certainly true that the capacities that dispose a thing to act in accord with its nature must be acquired through repeated acts. But even if this is the usual course of things, it is not at all obvious that this *must* be the case for every virtue. It is not at all obvious, that is to say, that the capacity to act in accord with a nature could not be divinely given.[45]

The response outlined in the preceding paragraph might, however, give rise to a still further objection. Even if one concedes that a habit might be divinely given, one might nonetheless argue that it would be irrelevant, at least where praise and blame are concerned. For when we say that an individual has a

---

[43] See Hursthouse 1999: 159; Zagzebski 1996: 118–121; and Roberts 1984: 235.

[44] Aristotle 1999: book 1 ch.7.

[45] Rosalind Hursthouse is willing to concede that virtues might in some exceptional cases be divinely given, but Linda Zagzebski's argument would seem to exclude even virtue as a divine gift, for she argues that it would be *logically impossible* for an individual to possess a virtue he had not acquired. I do not think Zagzebski's argument is successful, but any attempt to accommodate Aquinas's account would certainly need to respond to her objection. See Hursthouse 1999: 159; and Zagzebski 1996: 118–121.

virtue, we indicate that he is worthy of praise, just as we indicate that he is worthy of blame when we call him vicious. But praise and blame, in turn, are most comprehensible when the individual in question has actually *caused* the virtue (or vice) through his own repeated actions. Robert Roberts, for instance, argues against the possibility of "giving" an individual a moral identity on precisely these grounds. Even if it were possible, by some drug or injection, to make someone indistinguishable from a saint, Roberts argues, "[T]he praise for his saintliness and thus for his deeds, would not be *due* him."[46] If all of the essential features of Christian practical wisdom ultimately stem from God's free gift, then how can it be called a virtue at all?

This second objection is, I think, answered more easily than the first, and the answer points to the most important contribution that Aquinas's account can make to contemporary Christian virtue ethics. Virtues are so called because they make things good specimens of their kind. In the Christian account, the Christian virtues make things capable of sharing in the divine life. At the same time, however, it is essential to the Christian account that one cannot and should not take *credit* for this capacity. It is in an important sense central to any Christian account of virtue that the praise for one's virtues be due more to God than to the individual in question. Aquinas's account, I think, provides an apt reminder of this.

Assuming the challenges mentioned above can be met, Aquinas's insights indicate an exciting new direction for contemporary scholarship on Christian virtue. Acknowledging the divine contribution to and role in the Christian moral life would make contemporary accounts of Christian virtue more intuitive.[47]

References

Annas, Julia. (2005). "Virtue Ethics." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 515–536. Oxford: Oxford University Press.

Aquinas, Thomas. (1953). *On Truth*. Vol. 2. Translated by James McGlynn. Chicago: Henry Regency Company.

——— . (1964). *Summa Theologiae*. Edited by Thomas Gilby. Cambridge: Cambridge University Press.

———. (2005). *Disputed Questions on the Virtues*. Edited by E. M. Atkins and Thomas Williams. Cambridge: Cambridge University Press.

Aristotle. (1999). *Nichomachean Ethics*. Translated by Terrence Irwin. Indianapolis, IN: Hackett Publishing Company.

[46] Roberts 1984: 235.

[47] My thanks to Christian Miller for comments on drafts of this chapter. This chapter was made possible through the support of the Templeton World Charity Foundation. The opinions expressed in this chapter are those of the author and do not necessarily reflect the views of the Foundation.

Bejczy, Istvan. (2005). "The Problem of Natural Virtue." In *Virtue and Ethics in the Twelfth Century*, edited by Istvan Bejczy and Richard Newhauser, 133–154. Leiden: Brill.

Bernanos, Georges. (2002). *Diary of a Country Priest*. Boston: Da Capo Press.

Copelston, Frederick. (1962). *A History of Philosophy*. Vol. 2. New York: Image Books.

Foot, Philippa. (1997). "Virtues and Vices." In *Virtue Ethics*, edited by Roger Crisp and Michael Slote, 163–177. Oxford: Oxford University Press.

Garrigou-Lagrange, Reginald. (1939). *The Three Ages of the Interior Life*. Translated by Timothea Doyle. Paris: Éditions du Cerf.

Holy Bible. (1963). New York: Catholic Book Publishing Company.

Hursthouse, Rosalind. (1999). *Virtue Ethics*. Oxford: Oxford University Press.

Louf, André. (1992). *Tuning in to Grace*. Kalamazoo: Cistercian Publications.

Porter, Jean. (1992). "The Subversion of Virtue: Acquired and Infused Virtues in the *Summa Theologiae*." *Annual of the Society of Christian Ethics* 12: 19–41.

Roberts, Robert. (1984). "Will Power and the Virtues." *Philosophical Review* 93: 227–247.

Russell, Daniel (2008). "Agent-Based Virtue Ethics and the Fundamentality of Virtue." *American Philosophical Quarterly* 45: 329–347.

———. (2009). *Practical Intelligence and the Virtues*. Oxford: Oxford University Press.

Slote, Michael. (1997). "Agent-Based Virtue Ethics." In *Virtue Ethics*, edited by Roger Crisp and Michael Slote, 163–177. Oxford: Oxford University Press.

Stump, Eleonore. (2011). "The Non-Aristotelian Character of Aquinas's Ethics: Aquinas on the Passions." *Faith and Philosophy* 28: 29–43.

Vogler, Candice. (2012). "Aristotle, Aquinas, Anscombe and the New Virtue Ethics." Unpublished.

Zagzebski, Linda. (1996). *Virtues of the Mind*. Cambridge: Cambridge University Press.

17 }

Bearing Burdens and the Character of God in the Hebrew Bible

Cristian Mihut

A popular conception of God's character is neatly captured in Michael L. Morgan's summary of Judaic thought:

> The biblical God is a lawgiver. Sin is transgression of divine law: it is disobedience to divine command and an affront to God, an act of rebellion. The normative response of God to sin is anger and retribution. The biblical God is a just God, and this justice is a feature of divine life and cosmic order that is canonized in the book of Deuteronomy and expressed extensively throughout the Bible.[1]

Let us refer to this as the classical picture of divine character. If God is a legislator and justice has pride of place, what should we say of divine forgiveness? How does forgiveness fit within the classical picture? Some in the Rabbinic and Christian traditions think forgiveness is, at most, a release valve for divine wrath.[2] God necessarily has the disposition to forgive, they opine, but its characteristic function is to temper retribution. According to this view, forgiveness belongs to a framework of mainly forensic concepts; sin is a crime against the cosmic moral order set up by a punctilious judge who is offended by it, legislates against it, and fights it. Notice how even on Morgan's understanding, forgiveness seems parasitic upon and subordinate to divine forensic justice:

> Whether or not this compassion results in a reduction of punishment or stay of execution, as it were, there is more to God's response, and

[1] Morgan 2012: 139.

[2] For a terrific summary of the plurality of positions held by the Rabbis on the relationship between divine justice and forgiveness, see Urbach 1975. For a sense of how the Christian tradition articulated what I called the classical view, see Edwards 1741. Also, one hears echoes of this in the Westminster and the Heidelberg Catechisms.

> this supplement is what forgiveness involves . . . Forgiving is a surplus that includes a change of attitude, a sense of goodwill, and an overcoming of the sense of being violated, humiliated, or diminished. The Bible does not thematize this surplus, but it does hint in its direction.[3]

In Morgan's subtle appropriation of the classical view, forgiveness is a divine disposition that may soften or delay punishment while necessarily involving a specific change in God's self-regarding and other-regarding attitudes. Divine justice has theological, normative, and motivational primacy. Forgiveness is, at best, a tamer of retribution.

I argue here, however, that the narratives of the Hebrew Scriptures (what Christians regard as the Old Testament) suggest a more ambitious view of divine forgiveness and a corrective to the classical model. Forgiveness is not an anomaly at the outskirts of a fully integrated retributive deity. It is not chiefly the tamer of fierce justice, and not the therapy God needs to get over the sense of being violated. Morgan is correct that forgiveness implies a stable goodwill toward offenders. But this is not some "surplus" leftover after justice has debited most of God's resources. Forgivingness—the settled disposition to forgive—is at the center of God's agency, and is grounded in proactive, other-directed dispositions that figure deeply in the explanation of divine action.[4] And the Bible does thematize it. To paraphrase Morgan, we might very well say, *The biblical God is a hopeless forgiver. Sin is a burden that weighs down and diminishes offenders and their communities. God's normative response to burdened humans is primarily mercy and a readiness to carry their load*.

My main aim is to conceptualize an aspect of divine forgivingness, starting from the image of God bearing away Israel's iniquity. I first unpack the image of bearing away iniquity (*nasa awon*) and give four reasons for its centrality in the Hebrew Bible—its idiomatic frequency, paradigmatic frequency, liturgical depth, and fruitful conceptual overtones. After a methodological interlude, I develop a preliminary account of *nasa*-based divine forgivingness that involves two commitments: the commitment to remove the offense from the transgressor, and the commitment to absorb the consequences of the offense. Following the account, I illustrate how God absorbs rebellion in the story of Jonah. I end the chapter by addressing two main objections to this understanding of divine forgivingness.

---

[3] Morgan 2012: 142.

[4] Whenever I use "forgivingness," it refers to the entrenched disposition to forgive. Forgivingness is conceptually distinct from the act, the process, or even the non-entrenched disposition to forgive. See Roberts 1995.

I. Unpacking the Image

Before I focus specifically on the divine trait of forgivingness, I investigate more generally some ways in which the Scripture imagines divine forgiving actions. The argument here is predicated on the idea that the biblical metaphors that communicate the idea of divine forgiveness partially constitute the content of the concept to which they point. We begin from biblical images to order to enrich our concept of forgiveness, in the hope that a suitably enriched concept picks out novel and relevant aspects of the divine trait of forgivingness. But the images from which we start do not merely adorn the cognitive content of the concept; they enter the very kernel. And there are at least three central biblical pictures of divine forgiving actions. God is said to cancel the debt of our sins[5] (and the conceptually related, blot out iniquity),[6] to cover it (*kipper*),[7] and to bear away sin (*nasa awon*).[8] Before I go on to motivate the importance of the latter image, I make some preliminary genealogical and contrastive remarks. Though ancient, these metaphors are still with us, each in their own way motivating distinct (and sometimes competing) sets of intuitions that seek articulation in various contemporary forgiveness models. Additionally, the images of cancellation and of covering provide pertinent contrasts to the aspects of the concept of forgiveness derived from the image of lifting up iniquity, the image whose merits I want to highlight.

The first image suggests that forgiveness involves the annulment of an economic or legal obligation. Famously, the prophet Isaiah imagines the sin of Israel as a huge national debt accrued toward God, and the suffering in the Babylonian exile as payment for it.[9] All the same, God alone, as the wronged party, has the unique power to cancel the debt or to forego litigation: "I, I am He who blots your transgression for my own sake, and I will not remember your sins."[10] Having considered the philosophical extensions of this metaphor,

---

[5] This picture of forgiveness becomes much more prominent in the New Testament. See Matthew 6:12. Gary Anderson claims, for instance, that "in the New Testament the metaphor of sin as debt was ubiquitous" (2009: 31).

[6] Thus Psalm 51:9: "Blot out all my guilt."

[7] See Psalm 85:2–4: "You forgave (*nasa*) the iniquity of Your people; You covered all their sin. You withdrew all Your fury; You turned away from Your burning anger." Interestingly here, two images of forgiveness—carrying burdens and covering iniquity—occur side by side.

[8] See Exodus 34:8.

[9] Here Isaiah says:

> Comfort, O comfort my people, says your God
> Speak tenderly to Jerusalem,
> And cry to her
> that she has served her term,
> that her penalty is paid,
> that she has received from the Lord's hand
> double for all her sins. (Isaiah 40:1–2, New Revised Standard Version)

[10] Isaiah 43: 25 (NRSV).

some suggest that it comes dangerously close to conceptualizing forgiveness as forgetfulness. Avishai Margalit argues, for instance, that when faced with a choice between *blotting out* and *crossing out* iniquity, the latter is conceptually, psychologically, and morally preferable to the former. When the victim blots out or erases the moral debt accrued against her, she commits to forgetting the wrongdoing. On the other hand, crossing it out "leaves traces of error under the crossing-out line."[11] It seems preferable for victims to disregard the offense without forgetting it, as it is psychologically more realistic to aim at overcoming negative attitudes toward offenders than to pretend the offense did not happen. If these observations hit the mark, it might be hard to see how divine forgiveness understood as forgetting iniquity could be exemplary or paradigmatic for us. Perhaps it is coherent and morally worthy for God to cancel debts of sin and to blot out human iniquities. But it is unclear whether mere mortals have analogous powers at the interpersonal level.

Unlike the first image which evokes a legal and economic backdrop, the context for the second metaphor is cultic, rooted in the expiation rite of the Day of Atonement.[12] Once a year the high priest enters the holy of holies and purifies the holy space and objects therein by sprinkling them, covering them (from the Hebrew verb *kipper*) with the blood of a sacrificial goat. Because sin is conceived in the rite as a pollutant that desecrates holy spaces and the objects contained in them, transgression needs to be covered up or expunged for the purpose of safeguarding the continued presence of God in the midst of his people.[13] As I just noted, Avishai Margalit finds this biblical metaphor conceptually, psychologically, and morally promising. Covering iniquity, he says, implies crossing instead of deleting, and this suggests a model of forgiveness where the victim countenances the offense while forgoing the pursuit of retaliatory measures. In forgiveness I continue *to believe* that the offense puts a blotch on the personal and moral history of my offender, but I do not allow that stain to enter my practical deliberations about this person.[14]

---

[11] Margalit 2002: 197.

[12] See primarily Leviticus 16.

[13] The New Testament improvises on this trope in interesting ways. In the letter to the Hebrews, for instance, Jesus functions simultaneously as the high priest and as the sacrificial goat. He takes his own blood into the spiritual temple, constituted in part by the community of believers, and covers up their consciences making them fit to stand in the presence of God. (See Hebrews 9:11–28.)

[14] Nicholas Wolterstorff's recent account of forgiveness also relies on this metaphor, sometimes implicitly and sometimes explicitly. In his view, forgiveness is "the enacted resolution of the victim no longer to hold against the wrongdoer what he did to one—or to use biblical language, no longer to count the sin against him." (Wolterstorff 2011: 169). On Wolterstorff's view, to victimize another puts a blot on one's personal and moral history (where moral history is "that ensemble of things he did that contribute to determining his moral condition"). When we excuse the offender, we no longer believe that his offense is a blot upon his moral history. Importantly distinct from excusing, in forgiving we no longer hold the offense against the wrongdoer in the sense that *we treat him as if* the offense put a blot on his personal history alone but not on his moral history, although *we also believe* that the blot remains part of his moral history. The blot is covered but not erased; the transgression is remembered but no longer effectual in the practical deliberations that involve the offender.

Contrasting these two images, the third is primarily curative while also deeply rooted in Israel's cultic life. If the first two metaphors see God either as primarily a cosmic judge or as primarily the pure, Holy Other, the third casts God chiefly as a healer. God is mainly in the business of healing and unburdening a broken people. I flesh out other differences between bearing burdens and covering stains later on, but for now I want to make explicit my motivations for starting from *nasa awon*. I claim that the image of God carrying away iniquity reveals neglected but essential aspects of the biblical concept of forgiveness. And because of this, I claim that bearing burdens is important for conceptualizing divine forgivingness. I offer four mains reasons for the centrality of this image.

First, there is *idiomatic frequency*. Other idioms, naturally enough, convey forgiveness in the Hebrew Bible. But *nasa awon* is used more frequently than any other image of forgiveness in the Hebrew Bible. Too often English translations rendering it as "forgiveness" obscure the literal Hebraic idiom of "carrying burdens away." According to Gary Anderson, the most common way of expressing human sinfulness in the First Temple period is in terms of a burden or a weight, and the idiom of *bearing away a sin* "predominates over its nearest competitor by more than six to one."[15]

Second, there is what I call *paradigmatic frequency*. The most intimate portrait of God in the Hebrew Bible is revealed at Mount Sinai. Here Yahweh claims Israel as God's own. Here Israel falls for another deity, the proverbial golden calf. Here Moses asks to see God's face in the aftermath of something akin to marital betrayal. And as God shows up robed in divine glory, Moses hears the declaration of divine attributes, "God, showing-mercy, showing-favor [grace], long-suffering in anger, abundant in loyalty and faithfulness [truthfulness], keeping loyalty to the thousandth (generation), bearing iniquity, rebellion and sin, yet not clearing, clearing (the guilty), calling-to-account the iniquity of the fathers upon sons and upon sons' sons, to the third and fourth (generation)."[16]

Biblical scholars refer to this list of divine self-attributions as the "grace formula."[17] Notably, *nasa awon* makes the list. God's readiness to bear Israel's rebellion and guilt is part of God's keenest self-disclosure. And, crucially, the grace formula is repeated throughout the Hebrew Bible, in the law and the psalms, in the major and minor prophets.[18] In fact, the refrain of coordinated

[15] Anderson 2009: 17. Anderson finds 108 instances of *nasa*, 17 instances of *salakh*, and 6 instances of *kipper*.

[16] Exodus 34:6–7 in the translation of Fox 1997: 455.

[17] About the provenance of the notion of grace formula, see Wöhrle 2009: 6–7, n9.

[18] See the various iterations and interesting developments of this refrain in the law: Numbers 14:17–19; in the psalms: Psalms 103:6–14; 145:8–9; 86:5; 85:2–4; in the major prophets: Isaiah 63:7–9; and in the minor prophets: Joel 2:12–14; Jonah 3:9; 4:2; Nahum 1:2b, 3a; Malachi 1:9a; Hosiah 4:10; Joel 4:21; Micah 7:18–20.

attributes structures the Hebraic self-understanding and their understanding of the kind of deity that Yahweh is.

God's readiness to bear iniquity, alongside divine faithfulness, mercy, and slowness to anger are introduced at *the* foundational moment in Israel's history as a nation. Preeminent among God's attitudes and paradigmatically revelatory of his character, these explain divine action in the context of Israel's original betrayal. Scripted into the original national rebellion is the revelation of Yahweh as singularly faithful, compassionate, and forgiving. For genealogical reasons, these dispositions have pride of place in the characterization of divine agency.

In various cultures and traditions some narratives are revelatory of the distinctive moral excellences of these cultures.[19] As Robert C. Roberts notes, some paradigmatic cultural narratives become not just illustrative but constitutive of the self-understanding and moral identity of a people.[20] Roberts illustrates this point by reflecting on the way the Hebraic moral identity is constituted by the Exodus story:

> If an ordinary narrative displays the self-understanding of its characters, this one structures a self-understanding in its believing reader or hearer. This narrative does not just exemplify a moral tradition, in the way that, say, the characters of Alyosha and Father Zossima in Dostoevsky's *The Brothers Karamazov* exemplify Russian Orthodox Christianity and the virtues of its most exemplary adherents; it founds one or at least constitutes a crucial stage in its founding . . . The Hebrew sense of self that is involved in one way or another in all the Jewish virtues—hospitality to the foreigner, generosity, trust and humility, the praise of God, just treatment of fellows, contrition, truthfulness, marital fidelity, honoring of parents, and so forth—is in essential part the sense of belonging to the people whom God made his own in this special way.[21]

I want to extend Roberts's point. Clearly, the Exodus story, God's claiming Israel as his at Sinai, and the subsequent rebellion are, if not *the* constitutional events in the Hebraic self-consciousness, then surely among its most significant markers. But they are also foundational and paradigmatic episodes for understanding God's moral character. Theologically, it is not incidental that Israel abandons Yahweh immediately after God claims them at Sinai. And the dispositions the grace formula ascribes to God are no mere rhetorical flourish. Forgiveness, mercy, eternal faithfulness, and short-lived anger define essentially God's moral self in reference to fickle Israel. These are not disposable, contingent, or revisable qualities. They are *the* interpretative lenses that filter

---

[19] See among others, Roberts 2012; MacIntyre 1984; and Nussbaum 1990.

[20] Roberts 2012: 179–180.

[21] Ibid., 180.

Israel's subsequent history with Yahweh. The molten golden calf in the shadow of Sinai recalls the real story of the fall for the Hebrews. It is there that they betray Yahweh. But it is also there that God renews his vow to them, reminding them he has carried their burdens thus far, and binding himself to keep doing it. It is at Sinai that God forgives them. And when successions of rebellions, idolatries, and exiles threaten to unravel at once the sense of Yahweh's presence and the very fiber of Hebraic identity, Israel will deepen, sometimes in creative and innovative ways, the memory of a merciful and forgiving God.[22]

Third, forgiveness understood as God bearing away iniquity has *liturgical depth*. In the Hebraic mind, I claim, bearing away burdens refers to a nonnegotiable moral quality of Yahweh. It is an essential descriptor framing Israel's memory of an aspect of God's character. But *nasa awon* also plays a distinctive and irreducible role on the Day of Atonement (Yom Kippur), the yearly ritual enacted for the forgiveness of sins. And this distinctive ritual role reinforces in rich cognitive and emotive ways Israel's experience of divine forgiveness. That is to say, *nasa awon* has at once theological and existential depth.

I have already mentioned some theological resonances of *kipper* (covering iniquity), the cultic image of forgiveness associated with the first rite on the Day of Atonement, the sacrifice of the expiatory goat. But as is well known, the second rite famously involves a scapegoat. In the cultic action the high priest places the sins of the people upon its head, the scapegoat weighed down by the sins of the people being subsequently driven into the wilderness. The context here suggests that the metaphor "bearing away" is primarily curative.[23] If the expiation rite sees sin as a substantival miasma or a polluting smog "that creeps into God's presence . . . and wraps itself around the holy vessels," the scapegoating rite understands it as a concrete weight, a burden or a disease that affects people.[24] Covering sin evokes primarily a cultic context where sin is a contaminant and God is utterly holy. But the evocations of bearing away iniquity, while also finding a robust liturgical expression in Israel's cultic life, reveal God as the empathetic healer and deliverer of a bruised and burdened people. Thus Leviticus 16 goes out of its way to let us know that the cure is total: that the scapegoat carries away *all* iniquities, *all* transgressions, and *all* sin.[25] Developing this cultic theme of healing, Isaiah's faithful servant—the paradigmatic agent of God's deliverance for exiled Israel—is said to have "borne our infirmities and carried our diseases."[26] And Psalm 103 interestingly imagines *not the goat but God as the agent of removal*, though the image also clearly harks back

[22] See, in footnote 18, the litanies of texts that recall, amplify, and interestingly amend the initial grace formula.

[23] See Carmichael 2006.

[24] Hamilton 2005: 274.

[25] Leviticus 16:21.

[26] Isaiah 53:4.

to the scapegoat rite: "as far as the east is from the west, so far he has removed our transgressions from us." Some have perspicuously noted that in the larger context of this psalm the forensic idiom of forgiveness is grounded in the cultic and curative image of dispensing with iniquity.[27] God forgoes legal action against Israel *because* he is a God who heals and carries away iniquities.

From the liturgical context of the Day of Atonement, then, we can extract at least four main differences between sin as pollutant and sin as burden, differences that motivate radically different conceptualizations of the process and attitude of divine forgiveness. First, regarding the target object, pollutants affect the presence of God (they haunt holy places and infect holy things), but burdens affect people. Second, with regard to phenomenology, the subjective experience of sin as a burden *feels different* from experiencing sin as a stain. Third, regarding the nature and consequences of sin, non-intentional and casual wrongdoings were understood to result in impurities; by contrast, deliberate and wanton transgressions *burdened* communities and individual transgressors. Finally, regarding the vehicle and direction of sin removal, it is a consecrated and therefore pure animal that covers iniquity as its blood disinfects the spaces closer and closer to the heart of God's action in Israel: the holy of holies. But the scapegoat becomes impure and accursed as it hauls wantonness far from the holy of holies back to its source in the netherworld.[28] In light of these striking differences, one could make a case that *nasa awon* works at a more fundamental level than *kipper*. Instead of merely concealing successive accumulations of impurities that essentially remain there, the scapegoating rite effectively promises the decisive removal of the most egregious wrongs from the camp. But I do not need this conclusion. My argument does not require that *nasa* be more basic than *kipper*. All I need to show is that *bearing away sin* points to a distinctive and irreducible mechanism for expressing divine forgiveness on Yom Kippur. And the cultic evidence I presented bears this out.

I conclude this section with a final reason for starting from this metaphor. *Nasa awon* evokes fruitful semantic and conceptual connections to other proactive, favorable attitudes of Yahweh. The same verb used to express forgiveness is also used to describe the liberation of Hebrew slaves from the Egyptian captivity, where God carries them along "as on eagles' wings."[29] And a divine

---

[27] Gregory 2006: 549–551.

[28] Gary Anderson helpfully notes: "As Jacob Milgrom, the great scholar of the book of Leviticus, has argued, the goat that has been burdened with these sins 'is in reality returning evil to its source, the netherworld.' Because this area was thought to be beyond the reach of God, the sins would fall outside the range of his supervisory powers. Once God could no longer see them, it is as if they ceased to exist" (2009: 23).

[29] Exodus declares that at Mount Sinai the day before the giving of the Law, God tells Moses to impart this message to the whole people: "You yourselves have seen what I did to the Egyptians, and *how* I bore you on eagles' wings, and brought you to Myself. Now then, if you will indeed obey My voice and keep My covenant, then you shall be My own possession among all the peoples, for all the earth is Mine; and you shall be to Me a kingdom of priests and a holy nation" (19:4–6, New American Standard Bible). The same explicit image is remembered in Deuteronomy's poetic rendering of the

motivation for carrying Israel along that is highly thematized throughout the Exodus story and its recollections is divine empathy.[30] Yahweh just is the kind of God who experiences suffering from the perspective of the vulnerable in a way that has direct bearing on divine actions. As Isaiah 63:9, a text recalling at once the grace formula and the Exodus theme of divine empathy, expresses it: "So He became their Savior. In all their affliction He was afflicted, and the angel of His presence saved them; In His love and in His mercy He redeemed them, and He lifted them and carried them all the days of old."[31]

Interestingly, carrying Israel along animates not only divine empathy but also God's sympathy. God experiences the suffering of the helpless, not only from their perspective, but also from God's own perspective. God suffers not only alongside but also on behalf of the people. As I have already suggested, the transfer of the suffering due to transgression is clearly pictured on Yom Kippur. More generally, Terence Fretheim argues that representational suffering is interwoven pervasively in Israel's cultic and religious life.[32] The scapegoat, the priest, the prophet, and the suffering servant are all representatives of Israel that bear the suffering for the people.[33] And these representatives, especially the prophet and the suffering servant, paradigmatically embody the suffering of God *on behalf of the people*. God is thus moved to carry vulnerable Israel as a result of two complex and interlocking motivations: divine empathy and divine sympathy.

It is also worth noting a third resonance of *nasa awon*. The Hebrew word for glory *(kabod)* means literally weight or weightiness. Ancient kings would show their glory, their weight, bedecked in armor as they triumphantly entered the regal city. So when Moses asks God to reveal his glory at Sinai in the prequel to

---

wilderness experience: "He found him in a desert land, and in the howling waste of a wilderness; He encircled him, He cared for him, He guarded him as the pupil of His eye. Like an eagle that stirs up its nest, that hovers over its young, He spread His wings and caught them, He carried them on His pinions" (32:10–11, NASB).

[30] As many have noticed, God experientially knows the pains of Israel in Egypt (Exodus 2:23–25; 3:7–8); he "hears," "remembers," "sees," and "knows" their groaning pains. Also, see Judges 2:18: "And he saved them from the hand of their enemies all the days of the judge; for the Lord was moved to pity by their groaning because of those who afflicted and oppressed them."

[31] Some may worry here that the simple reference to God as the bearer of Israel's suffering is complicated by the term "angel of his presence." A couple of responses. First, given this particular translation (NASB), God seems to be the direct subject who experiences Israel's suffering, as well as who loves, shows mercy, and bears her. The angel of his presence appears as the direct subject of the act of redemption. Second, even if God's identity is a bit more fluidly connected to the angel of his presence, clearly the latter's agency is an expression of God's agency. That is, the angel does not perform actions that God himself would not perform. Finally, and most interestingly, apparently the original Hebrew is ambiguous. Alternative translations render it as a contrastive intensification: "It was no messenger or angel, but his presence that saved them" (NRSV). If this is the original intent, the acts of suffering for the love, mercy, redemption, and bearing of Israel are intrinsic and characteristic of who God takes himself to be.

[32] See especially chapters 9 and 10 in Fretheim 1984.

[33] He points to texts such as Exodus 28:28, Leviticus 10:17, and Ezekiel 4:4–6.

the grace formula of Exodus 33, he expects a dazzling display of martial majesty. What he gets instead is a list of divine attributes ending with a picture of God weighed down by sin and rebellion. Perhaps the narrator is subtly telling us that the weight of God distinctively involves the readiness to be weighed down by the afflictions of others, especially of the most vulnerable. And if the connection between divine glory and the fate of the vulnerable is merely suggested here, it is heavily developed elsewhere in the Hebrew Bible.[34]

II. Methodological Interlude

Before I expand the image into an account, I want to say a bit about methodology. The argument here belongs to a larger project of theological retrieval. I track down a genealogy and development of *one* key biblical metaphor of forgiveness and attempt to press it as far as it will "go" into an account of divine forgivingness. If one is interested in theologizing and philosophizing about biblical forgiveness or about the corresponding divine virtue, this is obviously not the only way to proceed. One approach distinct from mine would map the contours of the phenomenon of forgiveness by prioritizing modern moral intuitions and downplaying the relevance of ancient medical, ritualistic, or economic contexts. In the language inherited from Immanuel Kant, forgiveness proper would be understood as a kind of morally appropriate response to wrongdoing, a transaction between autonomous agents, but metaphors such as impurity, debt, or disease would be mostly idle.[35] A second approach would extract from the various ancient biblical pictures some structural or abstract features that would then be employed in a general theory. Unlike the first approach that either dismisses the metaphors or arbitrarily prioritizes one of them, the second approach simply homogenizes and flattens the images, reducing their content to the lowest common denominator.

Whatever virtues they have, be they simplicity or theoretical unification, these two alternative approaches also incur costs. The first unduly restricts the articulations of forgiveness within the contemporary framework of autonomous moral agency, when the phenomenon of forgiveness may have richer psychological, medical, or even cosmic resonances. Similar worries arise for the homogenizing approach. Different biblical metaphors of forgiveness presuppose different contexts—juridical, relational, psychological, medical, cultic, and economic—so a homogenous theory of divine forgivingness would tend to flatten richer descriptions and their distinctive functions in search of

[34] See, paradigmatically, Psalm 8.

[35] In my opinion, David Konstan comes close to holding just this kind of view. See Konstan 2010, especially chapter 1, "What Is Forgiveness," and chapter 4,"Divine Absolution: The Hebrew and Christian Bibles."

common ground. Though in many ways fruitful, these alternative approaches either dismiss the biblical metaphors or arbitrarily prioritize some at the expense of others, or compress their rich contents.

My particularistic approach prioritizes one metaphor, and does so for good reason. A more complete account would tell a story that somehow coordinates the remaining central biblical images. But just to be clear, I make no grand claims about theoretical completeness. Perhaps the biblical phenomenon of forgiveness is essentially untidy and amorphous, and the different biblical lenses we have at our disposal pick out at most a "cluster concept."[36] The concept may be resistant to a full analysis in terms of necessary and sufficient conditions, but could still be theorized by reference to various features that tend to cluster together around one or another biblical image of forgiveness.

My project mines only one image in the cluster, in the hope that it might be coordinated with the accounts derived from other images in the cluster. As we have just noted, the genealogy of forgiveness as covering iniquity (*kipper*) reveals that sin essentially attaches to and pollutes not people but the divine presence in Israel's camp. The genealogy of bearing burdens (*nasa*) shows that sin essentially burdens human beings. I do not know exactly how or whether these two pictures could be reconciled. But I think some really interesting theoretical results follow if one attends to the way in which the biblical understanding explores and expands each of these metaphors. I am doing this for bearing burdens, since I believe it is one of the more enduring and apt biblical pictures. Others may engage in a similar project, that of retrieving the metaphor of covering iniquity. And if some pictures turn out to be more apt and fruitful than others, we can also pursue coordination. But the task of theorizing a non-arbitrarily selected metaphor becomes foundational. This, then, amounts to a confession and a promise. I theorize in "the grip" of one of the better pictures of forgivingness.

III. From Image to Account

In section I, I argued that *nasa awon* is a frequent, paradigmatic, liturgically deep, and semantically resonant metaphor. We are now well positioned to garner the insights and characterize an important aspect of divine forgivingness. Observing its place in the grace formula and drawing upon the imagery of Yom Kippur, I first define the divine disposition to forgive primarily as a distinctive mode of God's *hesed* (divine faithfulness) that involves two interlocking commitments: God's commitment to distance individuals and communities from

[36] Mike Rea suggested to me in conversation this way of partially containing the ostensible untidiness of divine forgiveness.

their sins, and God's commitment to absorb the consequences of iniquity himself. Second, I contrast sympathetically this *nasa*-based conception of forgiveness with one derived from *kipper*.

I begin by noting that the place of *nasa awon* among the other divine names in the grace formula invites a network characterization of this divine trait. By network characterization I mean that forgivingness is elucidated in relation to other entrenched traits that constitute God's character. Though I sort out divine dispositions along cognitive, volitional, and affective dimensions, these simply reflect ways of culling helpful taxonomies for various divine attitudes. But the taxonomy should not be taken to carve out divine nature at its joints.

Wittgenstein argued that the grammar of a concept is given by its patterns of usage. We pick up the concept "slab" by noting how it functions in various linguistic contexts in which it is instantiated, such as when we speak of "a slab of meat," or "a slab of stone," or "a wooden slab." By analogy, says Robert C. Roberts, the grammar of a trait "which may vary from one moral tradition to another (Aristotelian liberality differs in grammar from Christian generosity, and both differ from the Nietzschean *schenkende Tugend* (bestowing virtue) . . .) determines such things as the kind of actions and emotions (*in situ*), of motives and reasons, that characteristically exemplify the trait."[37] *Nasa awon* accompanies in the grace formula other favorable and lively divine attitudes such as graciousness, loving-kindness (relational loyalty), compassion (mercy), long-suffering (slowness to anger), and truthfulness. These dispositions are significant, for they give us at least a provisional map of the divine moral character. Since forgivingness is introduced alongside these attitudes (and not others), they all point to a "grammar" of the trait. I propose, then, that the content and function of divine forgivingness is determined in part by reference to the divine cognitions, affections, and volitions recited in the grace formula.

Divine forgivingness is defined in part by reference to divine cognitions, such as truthfulness. This may be taken to involve a complete, compassionate, and undistorted knowledge of transgressors, of circumstances related to sin, of general and specific consequences of sin, and of God's own inner responses to sin and sinner. Divine truthfulness then is co-instantiated with, and part of the grammar of, forgivingness. God could not be forgiving without being truthful in the sense explained earlier.

Divine forgivingness is also accompanied by characteristic divine affections. Leaving aside a defense of the possibility of divine emotions, the relevant divine affections are mercy, anger, and long-suffering.[38] As the grace formula relates, the positive or favorable affections, such as mercy, have priority over

---

[37] Roberts 2012: 177.

[38] It may be surprising to some that "long-suffering" picks out an emotional disposition. Literally "long of pinions," the expression is almost always used in the Hebrew Bible to characterize feelings, specifically as a way of describing an emotion that counterbalances or slows down anger.

the reactive (negative) affections, such as anger and accusation. This priority can be cashed out in at least three distinct ways. First, favorable divine affections are more enduring than anger and accusation. There is an expiration date on anger, but not on mercy. Their endurance suggests that the favorable affections constitute the default setting of divine sensibility, while divine wrath is simply a momentary, first-level reaction to transgression. Second, mercy runs motivationally deeper than anger. This is true in virtue of a structural, non-contingent feature of divine agency. God's commitments and long-range plans tend to align with divine mercy over and against wrath. Finally, divine favorable affections contextualize and permeate divine anger itself. "Long-suffering in his anger" says the grace formula. Divine wrath is neither an expression of personal vengeance nor an objective, uninvolved disposition to dish out just deserts. Though God's anger flares, in paradigmatic biblical episodes divine grieving flanks its onset and divine regret and suffering its closure.[39] And if grieving and regret are aspects of compassion, paradigmatic divine anger seems framed, measured, and infused by mercy. Basing his judgment on the analysis of a number of "divine wrath" texts, Fretheim even goes as far as to claim, "lament is always an integral aspect of the wrath of God."[40]

We have seen, then, some of its cognitive and affective dimensions, but the essence of *nasa*-inspired divine forgivingness really lies in the divine capacity for maintaining commitments, what we may variably call divine volitions or decisions. It is essentially a distinctive aspect of divine *hesed* (read loving-kindness, covenantal faithfulness, or steadfast love). More than a mere affective inclination, *hesed* is a stable and permanent commitment to act lovingly toward Israel. It is indeed grounded in and shaped by divine mercy while, at the same time, lending support and reinforcing it. In a gripping iteration of the grace formula, the prophet Micah says,

> He does not retain His anger forever, because He delights in unchanging love [*hesed*]. He will again have compassion on us; He will tread our iniquities under foot. Yes, You will cast all their sins into the depths of the sea.[41]

Divine favorable affections sustain *hesed*, and vice versa, in ways that outstrip human notions of faithfulness to persons. It is of a different order and magnitude. God would be "within his rights" to break covenant with Israel when they are unfaithful. Yet, strikingly, time and again God is said to be

---

[39] Thus God's judgment in the Flood is framed at the beginning by divine grief: "The Lord was sorry that He had made man on the earth, and He was grieved in His heart" (Gen. 6:6, NASB). And at the dénouement, God seems to express a form of regret: "I will never again curse the ground on account of man, for the intent of man's heart is evil from his youth; and I will never again destroy every living thing, as I have done." (Gen. 8:21) Interestingly, in the renewal of God's covenant with Noah from Genesis 9, we may even see an intensification of divine suffering. God commits to carrying even more of the human burdens after his judgment. On divine regret, see van Wijk-Bos 2005: 234–242.

[40] Fretheim 1984: 110.

[41] Micah 7:18. (NASB).

everlastingly loving-kind and faithful, often in the contexts of Israel's recurring betrayals. And precisely because Israel seems to pathologically pursue affairs with sundry neighboring deities, God's *hesed* stretches beyond the comprehensible.

I am inclined to take *nasa*-based forgivingness as an expression of divine loving-kindness, because *hesed* fixes the contours of divine agency. In the jargon of Harry Frankfurt, *hesed* is volitionally necessary for God.[42] This means that loving-kindness fixes the boundaries, the limits of God's agency. Loving-kindness is a higher-order divine disposition that serves a regulative function in God's motivational profile. Yahweh wouldn't be Yahweh, were he to act in ways inconsistent with loving-kindness. As a complex person God can simultaneously be the subject of several clashing attitudes, such as anger, compassion, and grace. But God's agential profile is constituted essentially by loving-kindness. This is not, as it were, just another item, not even "chief among sisters" in the divine psychology. Rather, it designates the very shape of God's decisional profile, the control tower advising God to throw his weight behind some first-order attitudes rather than others. And given that God's moral center is essentially constituted by *hesed*, he is already predisposed to identify with his favorable affections and, as it were, dial down his reactive (negative) affections. Because of his loving-kindness, God is the kind of agent that identifies with mercy and alienates retribution.

There is good biblical evidence to draw an intimate link between *hesed* and forgivingness as bearing burdens. Consider verses 11 and 12 from Psalm 103:

> For as the heavens are high above the earth, so great is his steadfast love toward those who fear him; as far as the east is from the west, so far he removes our transgressions from us.

I have already noted that in the larger context of the psalm the forensic idiom of forgiveness is grounded in the cultic image of dispensing with iniquity.[43] God forgoes legal action against Israel *because* he is a God who carries iniquities. Here I want to emphasize the poetic parallelism that expands the notion of divine faithfulness into the notion of forgivingness. The vertical expansiveness of steadfast love maps onto the horizontal distance God sets between people and their sins. The poetic device urges us to take the possibility of an internal connection between the two divine dispositions seriously: either the two are identical or forgivingness is an intensification of *hesed*, or, both are different expressions of the same underlying divine commitment or, at the very least, though distinct, the two have analogous functions.

In any case, it seems we have good reason to understand divine forgivingness as intimately connected to *hesed*, a higher-order disposition that in part

---

[42] See, especially, Frankfurt 1998, 2004.

[43] Gregory 2006: 549–551.

God identifies with his mercy over against his anger. Can we say more about the content of this divine steadfastness? I think we can, especially when we unpack the thick descriptions inherent in the scapegoat rite on Yom Kippur. The rite implies at least two things: that God sets a distance between people and their sins, so that now they are *de facto* unburdened, and that God himself is now burdened with the consequences of sinfulness. I suggest then that the distinctive content of *nasa*-based forgivingness is given by two interlocking commitments: (Ci) God's commitment to set a distance between individuals/community and their sins, and (Cii) God's commitment to absorb in himself the consequences of iniquity.

That forgivingness implies something like (Ci) is generally, if not universally, accepted. It is hard to see how an individual is forgiving if she does not regard her offenders in some ways as detachable from their offenses. As I noted, the Bible celebrates a God inclined to see sinners in a sympathetic and empathetic light. Given God's default affections, he prioritizes the circumstances that enable him to construe offenses as peripheral to the agency of offenders, and sinners as subjects of continual divine goodwill, even when the offense is engrained in their character. As we will see in a moment, God maintains his goodwill toward Jonah, even though the latter displays nontrivial, consistent, and entrenched patterns of rebellion against God.

Interestingly, *kipper*, the metaphor of covering sin, implies a specific expression of (Ci). *Kipper* indicates that the offense is disregarded without being forgotten.[44] God covers the shame of Adam and Eve with animal skins, and the guilt of Cain with a mark intended to protect him from being killed. As these paradigmatic stories reveal, by covering their shame and guilt God construes these first offenders as continued recipients of divine benevolence. But the cover itself draws attention to their offending status. And so the remaining traces of transgression require the exile of the agent. These stories suggest then that *kipper* is part of a conceptual network that essentially involves exile and repentance. And the interpretative tradition has inferred that only repentance or remorse affords a "non-magical way of undoing the past."[45] *Kipper* implies that God's commitment to fully separate us from our sin depends upon our commitment to separate ourselves from it.

But if (Cii) informs (Ci), then *nasa awon* suggests a more ambitious conception of divine forgiveness than *kipper*. Though human remorse may generally trigger divine forgiveness, to claim that the latter depends on the former is an overreach. Making divine forgivingness dependent upon the thinness, double-mindedness, and rationalization that attend (with psychological regularity) human repentance seems to me inadequate. If anything, the dependence

[44] See the discussion in Margalit 2002: 188–200.
[45] Ibid., 194.

relation runs the other way. We can only repent and lament our sinfulness due to God's robust commitment to separate us from it. Even our best and most sustained efforts of repentance are but feeble responses to this prior divine commitment or conflicted participations in it.

Additionally, if divine lament for our recidivist and rebellious condition shapes divine sensibility, God remains committed to his people even in the absence of occurent or dispositional remorse. Hosea 11 grippingly illustrates this. Pictured as a loving mother, Yahweh recalls teaching an infant Israel to walk, healing him when he stumbles, and nurturing him. But now he is an obstinate and deviant teenager, "bent on turning away from me." For a moment God entertains the notion to exile Israel back to Egypt and Assyria, lands that typify disintegration of identity. But that's just a passing thought. Yahweh is defined by reference to divine *hesed*. As a result of this basic commitment to an unremorseful child, itself *an expression of God's holiness*, Yahweh absorbs into his own being the consequences of Israel's disobedience ("my heart recoils within me"), while concomitantly outpouring his compassion and suppressing judicial action:

> How can I give you up, Ephraim?
>
> How can I hand you over, O Israel?
>
> How can I make you like Admah?
>
> How can I treat you like Zeboiim?
>
> My heart recoils within me; my compassion grows warm and tender. I will not execute my fierce anger; I will not again destroy Ephraim; for I am God and no mortal, the Holy One in your midst, and I will not come in wrath.[46]

Kipper then offers this paradigmatic understanding of forgiveness: God commits to set a distance between himself and offenders, to conceal their shame and guilt, but also to have them live through the consequences of their own sin. *Nasa awon* goes much beyond this conception to imply that offenders no longer have to carry the burden of their iniquity, and that God himself will live through the (remaining) consequences of *their* sins. While there is overlap between the two conceptions, *nasa awon* has a distinctive content that suggests a radical divine identification with the condition of the sinners. God is committed to absorb their pain and the consequences of separation, even though they will not repent.

In this section I have proposed an account of divine forgivingness derived from the picture of *nasa awon*. As a higher-order divine commitment, as a volitional necessity, it is a mode of *hesed*, a particular expression of God's faithfulness to his people. But even as a second-order commitment to identify with mercy and sideline anger, it has a specific content. This content is

---

[46] Hosea 11:8–9 (NRSV).

specified by the two interlocking commitments expressed in (Ci) and (Cii). Partially grounded in divine favorable affections, the two commitments of forgiveness in turn reinforce mercy as a deep-seated divine attitude. Forgivingness never lessens God's truthfulness, nor does it have to eradicate anger. For given the contours of divine *hesed*, truthfulness is wedded to divine compassion, and anger to divine lament.

IV. Bearing the Burdens of Jonah

In sharp contrast to the traditional picture of an inflexible cosmic judge, we have a vivid illustration of divine forgivingness in the book of Jonah. God is actively (not just reactively) involved in bearing away the rebellion of Jonah and of the pagans in the story. We meet Jonah as he scuttles away from God's command. But then his retreat into the belly of the ship, the belly of the fish, and the shadowy booth discloses an ever persistent and expansive willful separation from the concerns of a compassionate God and of needy others. His only communications with the pagans in the story are brief sentences proclaiming first self-doom, and then impending doom for them. And in his dialogue with God, Jonah oscillates between an accusatory anger and a defensive self-pity. Significantly, the prophet explains God's forgiving the Assyrians by reference to the memorable grace formula. But as others have noticed, Jonah does not celebrate or strive to model it. He accuses God for his forgivingness!

Jonah grasps that God is essentially forgiving, gracious, faithful, and compassionate toward Israel. But in his mind, the divine *hesed* toward Israel demands punitive measures against those devastators, the Assyrians. The God who relaxes his exacting justice either does not properly attend to the ruin Assyrians brought on Israel, or knows it and does not care, in which case God fails in faithfulness to Israel. In either case, God has given up on truth, and Jonah will bring it back into divine focus through any means available, be they disobedience, disaffection, minimal conformance, or tantrums. Compassion to the unforgiveable, Jonah thinks, necessarily violates the truth of the victim. Behind his willful blindness, underlying his flight away from God and others, and underneath his general grumpiness and expressed death wish, there is a gashing wound that, as far as Jonah can see, demands a strict retributive god.

Central to the story, however, is a divine commitment to both repentant violator and unrepentant victim, to both the contrite outsider and the callous insider.[47] And this divine forgivingness, though motivated by grace and

[47] Some may wonder if I am not obscuring the most important category, that of unrepentant violator. I do not think so. Jonah fits the bill. In the story he clearly represents the Hebrew victim wounded by Assyrian violence. But he is also an unremorseful and repeated violator of divine commands.

compassion in both cases, is unpacked in distinctive ways. Jonah had come to Nineveh wanting the divine fire to consume the pagans. Instead, divine compassion arranges the conditions God knows would elicit God's forgiveness. Notice that God has all the initiative in the story. God sends the Hebrew prophet to the Assyrians, he intervenes to reroute him when Jonah goes off course, and then resends him to the same destination. Jonah himself recognizes God's grace as the primary motive for divine action toward Nineveh: "for I knew that You are a gracious and compassionate God, slow to anger and abundant in lovingkindness, and one who relents concerning calamity."[48] And it is precisely this knowledge of God's gracious commitment that flings the prophet into throes of anger and despair. First Jonah accuses God of untruth and a lack of loyalty, and then, after losing the plant that God had providentially grown as shade for him, Jonah simply reports anger-unto-death. How does God respond?

> You had compassion on the plant for which you did not work and *which* you did not cause to grow, which came up overnight and perished overnight. Should I not have compassion on Nineveh, the great city in which there are more than 120,000 persons who do not know the difference between their right and left hand, as well as many animals?[49]

The priority of mercy is a metaphysical truth deeper than the truth Jonah thought God needed to relearn. Recalling the creation story of Genesis 2 where, as an artisan attached to his masterpieces, God crafts out of the dust *all* living creatures, God here includes not only the Assyrians but also their cattle in the range of his faithfulness and empathy. If God is moved by the imminent doom of beasts (parts of nature), he will spring into action at the impending plight of ignorant people. The God attending to the movement of blistering winds, leafy plants, gnawing worms, swallowing fish, and many cattle, is especially moved by the plight of people.

More subtly, God invites the prophet to countenance a particular expression of divine commitment toward him, a forgivingness that subverts Jonah's retributive demands. Running to Tarshish, Jonah shows he does not know his left hand from his right. No different from the Assyrians, he is the beneficiary of gratuitous grace. God spoke to Nineveh only through a reluctant, tight-lipped, doom-and-gloom preacher. But look at all the messengers God had sent Jonah: God's word, the sea storm, the big fish, the plant, the worm, and the blustering wind. And to no avail. Jonah persists and entrenches himself in callousness. If even a weaker version of Jonah's truth applies, if even a weak retributivism rules the day, divine punishment is the only appropriate

---

[48] Jonah 4:2 (NASB)

[49] Jonah 4:10–11 (NASB). The italicized emphasis is in the translation.

response to willful, persistent rebellion. But Jonah's truth is not God's truth. God gives long-lived grace to those wearied and diminished by wounds. We might say, he puts up with them, entertains their questions, and carries their burdens.

Finally, God's forgivingness plays to a common denominator in order to regain Jonah's trust. Though corroded by anger, Jonah is not beyond the ability to lament. And if lament is not lost, compassion can blossom still. By drawing attention to his compassion for the lost plant, God attempts to pull Jonah's sensibility from the brink of death. Jonah is resentful and peevish. God aims to restore in him a disposition that the two of them share in common. Jonah impugns divine majesty. But God does not rebuff. Instead, God bids Jonah to drop his retributivism regarding people *and* to build on his natural pity for the desiccated plant.

I have argued that starting from the central biblical image of carrying away burdens, we gain fresh insights into God's disposition to forgive. I have proposed that forgivingness is a deeply engrained trait of God, a mode of *hesed* involving two interlocking commitments to the well-being of perpetrators. And I showed that the book of Jonah portrays a God in whom the disposition to carry away iniquity runs deeper than and subverts retributive justice. I close by briefly discussing two main objections to the picture I have painted.

V. Two Objections

1. IS GOD A MORAL MONSTER?

I claimed the Hebrew Scriptures speak of a deeply forgiving God. How do I explain the so-called texts of terror, the litany of terrifying episodes that seem to portray a god "not merely unworthy of worship, but deserving of moral censure"?[50] How about Sodom and Gomorrah incinerated; Dathan, Abiram, and their families swallowed by the desert; "the men and maidens of Midian, the tortured soul of Isaac, the daughter and son of Jepthah and David, upon who memory has bestowed no names, the Canaanite babes laid waste in war"?[51] If in speech or act the biblical God is directly responsible for genocides and atrocities, we do not even have rational grounds for divine forensic justice let alone forgivingness. The Scriptures are either an unreliable source of moral knowledge, or there is radical discontinuity between divine and human normative concerns. If the Scriptures manifestly track moral truths and there is continuity between human and divine moral intuitions, it would seem that some divine actions are horrific. And if God is not responsible for atrocities, at

[50] Fales 2011: 107–108.
[51] Ibid.: 108.

least I have to come to terms with the many and seemingly contradictory biblical descriptions of God's moral character, and provide a principled hermeneutic that de-emphasizes the texts of terror. Unifying God's character around forgivingness pays the price of arbitrarily sidelining litanies of atrocities and non-restorative retributions the Scripture ostensibly attributes to God. This, at least, is a version of the worry. And it's a big one.

To answer, I do not think God is morally corrupt, I emphasize the normative continuities between human and divine agency, and I see the whole Scriptures as a reliable guide to moral knowledge. Written by fallible humans embedded in specific cultural and historical contexts, often distinct from each other and clearly different from our own, the Scriptures predictably speak in various and not always concordant voices about God's character. In cultures and times where "the warrior" is a paradigm of excellence, it is morally permissible and perhaps practically necessary that God convey excellence through martial idiom. Consistent with the traits uncovered in the grace formula, God has long-range patience and graciousness even regarding incomplete human thinking patterns about him. All the same, if the whole of Scripture is superintended by a morally perfect being, then we should expect to find in it reliable ways to distill from the polyphony indices that point to increasingly more complete moral pictures of divinity. I have argued that the divine self-disclosure at Sinai is one such index. The prophetic recalibration of God in the aftermath of Babylon is yet another.[52] And for a Christian as myself, the New Testament accounts of the Incarnation of the Son of God provide yet another pivotal index.

I have argued that there are good genealogical and theological reasons to take God's self-attributions at Sinai as paradigmatic for revealing God's character. Israel is born there, defined and named by reference to the moral nature of its deity Yahweh. But for reasons having to do with a deepening of national self-consciousness, the exile provides another distilling moment, and a more finely tuned understanding of divine nature. Through the pivotal use of the grace formula, the book of Jonah sees Yahweh not only as a tribal god faithful to Israel alone, but also gracious and compassionate to all humanity. The idiom that originally demarcated Hebrew from gentile is now explicitly and perhaps subversively employed to showcase a cosmopolitan God proactively showing loyalty to the nations. And when looking retrospectively from Jonah

[52] I have already suggested that the book of Jonah develops the grace formula in novel and interesting ways, beyond the original focus on God's mercy on Israel's future. In Jonah, God's mercy and forgiveness has a global reach. Additionally, in some paradigmatic prophetic reiterations of the grace formula the original warnings about God's visiting punishment on Israel drop off entirely. See, for instance, how Hosea 11 suggests that God will not execute his anger, and will act instead on his compassion. Also, see Joel 2:12–14, where the prophet again invokes on God's mercy, loving-kindness, and compassion, and the divine visitation on iniquity disappears from the recounting.

onto the rest of Scripture, we can also see that there were hints of God's *hesed* to the nations all along. We see it in God's compassion to Hagar and in the redemption of Tamar, Rahab, and Ruth, all pagan women with questionable reputations and no futures. We see it in God prosecuting the injustice against Uriah the Hittite, God advocating for the ethnic minority over and against his beloved King David. But perhaps it is only now, post-Jonah, that these texts strike us as paradigmatically revealing of God's moral character.

Accordingly then, one part of my interpretive approach is a paradigmatic reading of the grace formula. But there are other moving parts. My hermeneutic is also particularistic and programmatic. It is particularistic in this way: there is no general recipe or solution to the texts of terror. We have to treat these case by case. Much of what is taken as the divine character is given to us through literature, and literature is not given easily to codification and generalization. The inferences we make about God's character from these Hebrew stories must be sensitive to conventions of genre as well as to the literature of the neighbors. If the flood story is staple ancient Near Eastern mythology, the focus surely is not on divine destruction of humanity (that's part of the assumed cultural context; all the gods are doing it). Instead, it is lament and restoration that is the distinctive Hebraic insight into the divine nature. Recall God's speech act to Abraham, interpreted by many to imply the command to sacrifice a son: "Take now your son, your only son, whom you love, Isaac, and go to the land of Moriah, and offer him there as a burnt offering on one of the mountains of which I will tell you."[53] If God attempts to jolt Abraham's dulled conscience for having too facilely sacrificed Ishmael his other son, then this divine speech act is not a universal moral law, notwithstanding the ostensible grammar. It is a call to confession and repentance. If the book of Judges details massive failures in practical judgment by the rulers, God is not blamable for Jepthah's vow to kill his daughter; rather, it is the human inability to discern the divine voice in a highly syncretistic culture. And if parts of Scripture seem to give expression to revenge fantasies or demonization of the enemy, it is more likely that these are staples of cultural modes of expression rather than of divine character.

This brings me to the programmatic aspect of interpretation. When I read the Scriptures I look to them to affirm what is most beautiful, just, and noble in human nature. Hope rekindles when I see how some later text, such as Jonah, expands, completes, or subverts the flood story. Delight deepens faith when I observe how the Lucan Jesus reconfigures the Joshua conquest narratives. But there may be biblical stories that resist retrospective completion or re-evaluation. Job may still seem the butt of a cosmic joke between God and Satan, and whatever intrinsic goods he gets by suffering can hardly be taken to justify the devastation wrought on his children. The point is that there

[53] Genesis 22:2.

are or might be stories like Job's that resist any explanation that rescues God's character. So what to do? Is it rational to give up belief in a good God? I propose that in light of the paradigmatic and particularistic interpretive approach I sketched, it is rational to hold these stubborn texts in tension with what I take to be main sweep of Scripture, but without giving up belief in a gracious and good God. It is all right to live with this tension. In fact, the temptation to resolve it too soon, one way or another, may be a sign of epistemic vice.[54]

2. DIVINE FORGIVINGNESS IS IMPOSSIBLE IN LIGHT OF GOD'S MORAL PERFECTION

God's moral nature consists of the sum total of all moral perfections. Perfection entails lack of any imperfection. But the ability to suffer is an imperfection. And the picture of divine forgivingness I painted insists that suffering for or suffering with are some of its preconditions. If God were forgiving in the sense I discuss, God would not be morally perfect. At least that's how the objection goes.[55]

It seems to me that the claim "the ability to suffer is an imperfection" must be earned in argument. We can no longer presume, as in the times of Plato, the Stoics, or Plotinus, that being unaffected is an excellence in or out of life. I, for one, remain convinced by Wolterstorff's contention that suffering is required in order to say a decisive "No!" to the particular desolations of our world and a firm "Yes!" to its splendors. In his compelling defense of divine suffering love, he argues against the view that elevates impassibility to divine status:

> So we are faced with a choice. If one adopts the Plotinian concept of God, the conclusion falls out that God does not suffer, and, of course, does not have passions. But one gets the conclusion by paying the price of removing from God all knowledge of, and love for, the particular things of this world. The question, then, is whether this price is too high. Virtually the entire Christian tradition would say it is. The Christian cannot surrender the conviction that God knows and loves his creation. Or to put it differently: to pay the price charged by the Plotinian concept of God is to move away from Christianity toward some other form of religion.[56]

I do not have much to add to Wolterstorff's claim. I simply note that the God celebrated by "virtually the entire Christian tradition" is taken by the same tradition to be the Sovereign Lord we meet in the Hebrew Bible. And this deity, if we recall the prologue of Exodus, defines Godself by reference to the ability to

---

[54] For a fantastic recent collection that addresses the problem of the moral character of God in the Hebrew Bible see Bergmann, Murray, and Rea 2011.

[55] A version of this objection is expressed in Minas 1975: 138–150.

[56] Wolterstorff 1988: 223.

experientially know the cries of the oppressed and by the multifaceted attachments he forges with particular families and individuals. Removing all this from God is too high a price to pay indeed.

It might be replied that the divine moral character is entailed by God's metaphysical nature, and divine suffering is simply incompatible with God's metaphysical nature as understood by classical theology. After all, how could a being that is at once immutable, omniscient, omnipotent, simple, and timeless also be affected by human concerns? While it is much beyond my project to assess each of these metaphysical properties of God in light of his capacity for suffering, I limit myself to making a methodological point.

It is important to examine, affirm, and internalize the features of God's metaphysical nature as understood by classical theology. However, the ancient stories recounting God's acts in history have a particular and singular normativity for believers. Saying that God is all powerful, all knowing, and immutable does not give us much content about what God is like. The Hebrew stories tell us more about this deity; particularly, we find out *that* he acts in history, *that* he is the Deliverer and Israel's King. But even these stories must be read in a particular light, with attention to the moral character of the agent doing the saving, to the particular texture of his personality, and to the bundle of motives moving him. As Fretheim puts it, the moral character of the divinity is revealed in the way he performs these actions:

> In short, it is not enough to say that God is the one who saves and blesses in these stories; what is crucial is *the kind of* God who is understood to be saving and blessing. A capricious God can save and bless. Even an impersonal God could engage in such activities. Is such divine activity motivated by love, or are there other factors? It has been demonstrated that there is nothing special about Israel's believing that God acts in its history, except that it may have done this more often and more consistently than other peoples. There is, moreover, no uniqueness to be found in the fact that its God was believed to be a saving and blessing God, though these are important words to preserve in any OT Theology.[57]

Fretheim can be taken to make two fundamental assumptions here. First, the uniqueness of the biblical God is primarily revealed in his moral character. Second, God's moral character is primarily disclosed through particularistic divine agency, that is, through the way in which God performs divine acts in Israel's history. These strike me as eminently compelling assumptions. They, in fact, should function as constraints on our biblical and philosophical theologizing. In light of this methodological turn, if a metaphysical account entails that God could not be long suffering, forgiving, compassionate, and gracious, we would have a *prima facie* reason to be suspicious of that picture.

---

[57] Fretheim 1984: 24.

But I want to end on hopeful and not on a suspicious note. I have argued here that we can develop an account of divine forgivingness starting from the frequent, liturgically rich, and theologically foundational metaphor of God's bearing away iniquity. I have also argued that divine forgivingness, understood as the interlocking of the joint commitments to remove transgression from persons and to absorb in God's own self the consequences of iniquity, is central in explaining key divine actions throughout the Scriptures. If forgivingness is a central feature of divine agency, then we have every reason to believe it must be consistent with God's metaphysical nature, whatever that turns out to be.[58]

References

Anderson, Gary. 2009. *Sin: A History*. New Haven, CT: Yale University Press.

Bergmann, Michael, Michael J. Murray, and Michael C. Rea, eds. 2011. *Divine Evil? The Moral Character of the God of Abraham*. Oxford: Oxford University Press.

Carmichael, Calum. 2006. *Illuminating Leviticus: A Study of Its Laws and Institutions in Light of the Biblical Narratives*. Baltimore, MD: Johns Hopkins University Press. 2006.

Edwards, Jonathan. 1741. "Sinners in the Hands of an Angry God. A Sermon Preached at Enfield, July 8th, 1741." In *Electronic Texts in American Studies*, paper 54, edited by ReinerSmolinski. http://digitalcommons.unl.edu/etas/54.

Fales, Evan. 2011. "Satanic Verses: Moral Chaos in Holy Writ." In *Divine Evil? The Moral Character of the God of Abraham*, edited by Michael Bergmann, Michael J. Murray, and Michael C. Rea, 91–108. Oxford: Oxford University Press.

Fox, Everett. 1997. *The Five Books of Moses*. New York: Schocken.

Frankfurt, Harry G. 1998. "Autonomy, Necessity and Love." In *Necessity, Volition, and Love*, Cambridge: Cambridge University Press. 129–141.

———. 2004. *The Reasons of Love*. Princeton, NJ: Princeton University Press.

Fretheim, Terence E. 1984. *The Suffering of God: An Old Testament Perspective*. Philadelphia: Fortress Press.

Gregory, Bradley C. 2006. "The Legal Background of the Metaphor for Forgiveness in Psalm CIII 12." *Vetus Testamentus* 54: 549–551.

Hamilton, Victor P. 2005. *Handbook on the Pentateuch: Genesis, Exodus, Numbers, Leviticus, Deuteronomy*. Grand Rapids, MI: Baker Academic.

Konstan, David. 2010. *Before Forgiveness: The Origins of a Moral Idea*. Cambridge: Cambridge University Press.

---

[58] Many thanks to Chris Brinks Rea, Mike Austin, Nancy Snow, David Dejong, Ruth Armstrong, Angela Knobel, and Jeff Bain-Conkin for helpful comments on earlier drafts of this article. I am especially grateful to Christian Miller for his careful, detailed, and thorough attention to earlier versions of this argument. This project was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

MacIntyre, Alasdair. 1984. *After Virtue: A Study in Moral Theory*. Notre Dame, IN: University of Notre Press.

Margalit, Avishai. 2002. *The Ethics of Memory*. Cambridge, MA: Harvard University Press.

Minas, Anne C. 1975. "God and Forgiveness." *Philosophical Quarterly* 25: 138–150.

Morgan, Michael L. 2012. "Mercy, Repentance, and Forgiveness in Ancient Judaism." In *Ancient Forgiveness: Classical, Judaic, and Christian*, edited by Charles Griswold and David Konstan, 137–157. Cambridge: Cambridge University Press

Nussbaum, Martha. 1990. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press.

Roberts, Robert C. 1995. "Forgivingness." *American Philosophical Quarterly* 32: 289–306.

———. 2012. "Narrative Ethics." *Philosophy Compass* 7: 174–182.

Urbach, Ephraime E. 1975. *The Sages*. Jerusalem: Magnes Press.

van Wijk-Bos, Johanna W.H. 2005. *Making Wise the Simple: The Torah in Christian Faith and Practice*. Grand Rapids, MI: Eerdmans.

Wolterstorff, Nicholas. 1988. "Suffering Love." In *Philosophy and the Christian Faith*, edited by Thomas V. Morris, 196–237. Notre Dame: University of Notre Dame Press.

———. 2011. *Justice in Love*. Grand Rapids, MI: Eerdmans.

Wöhrle, Jakob. 2009. "A Prophetic Reflection of Divine Forgiveness: The Integration of the Book of Jonah into the Book of the Twelve." *Journal of Hebrew Scriptures* 9: 1–17.

18 }

Domain Specificity in Self-Control

Angela Lee Duckworth and Eli Tsukayama

On March 17, 2008, Eliot Spitzer resigned as governor of New York. His rapid ascent up the political ladder had ended abruptly when it was discovered that he had been patronizing a high-priced prostitute ring. Spitzer's resume includes graduating Phi Beta Kappa from Princeton University and serving as an editor for the *Harvard Law Review*. By all accounts, he had a sound work ethic. Presumably, to succeed as an attorney and politician, Spitzer had to control his emotions when interacting with others. He did not smoke, do drugs, or drink heavily. Thus, it seems that Spitzer exhibited self-control in most domains. Yet, he behaved impulsively when it came to sex. How do we explain Eliot Spitzer? More broadly, why do some people act morally in some situations but not others? In particular, how do we reconcile apparent inconsistencies in self-control behavior? That is, why does it appear that an individual can be self-controlled in one situation (e.g., work) but impulsive in another (e.g., extramarital sex)?

In this chapter, we review the research on domain-specific impulsivity. We propose a theoretic model that can explain both domain-general and domain-specific impulsive behavior (and thus Eliot Spitzer's conduct), and discuss empirical studies that provide support for this model.[1] After discussing the relationship between self-control and character, we provide some historical background on the domain-specific versus domain-general issue. We then review the extant research and present some preliminary findings from a new longitudinal investigation of domain-specific impulsivity. We conclude with implications and future directions.

[1] We have discussed this theoretic model and some other ideas presented in this chapter previously (Tsukayama, Duckworth, & Kim, 2012).

Self-Control and Character

We define self-control as the ability to regulate behavior, thought, emotion, and attention in the service of valued goals and standards; and we define impulsivity as the lack of self-control.[2] As Hartshorne and May (1929) note, "[S]elf-control has been included among the virtues of the civilized man from the beginning of history. The writings of wise men, whether of the ancient cultures of China and India, the Nile valley and Asia Minor, or Greece, exalt industriousness and temperance, in evident fear lest the indolence and self-indulgence of the ordinary man prove his undoing" (p. 283). Moral character is, of course, multidimensional, but self-control appears to be a central virtue insofar as the ability to overcome impulsive responding for the sake of future consequences is necessary (though not sufficient) for the exercise of most other virtues (Baumeister, 2005). Blasi (2005, as cited in Lapsley & Narvaez, 2006, p. 265) proposed that "moral character is best described not by reference to lower-order virtues, such as honesty, generosity, and humility, among numerous others," but by higher-order virtues that include self-control (as well as integrity and moral desires, which Blasi also considered higher-order virtues).[3] Similarly, Baumeister (2005) explains, "[S]elf-control can be understood as the psyche's 'moral muscle' because it is useful for resisting temptation and performing other feats of moral choice. Indeed, self-control can be considered as underlying human virtue generally" (p. 65).

Historical Background on the Person-Situation Debate

Although it is now generally recognized that behavior exhibits both domain-general and domain-specific aspects (Epstein & O'Brien, 1985; Fleeson & Noftle, 2008), some historical background on the general versus specific debate (a.k.a., the person-situation debate: i.e., whether personality or the situation is more important in determining behavior) will help frame this review. By *domain general*, we mean that behavior that reflects a trait in one situation is highly correlated with behavior across different situations. For example, impulsive behavior in the food domain correlates strongly with impulsive behavior in the alcohol domain. We also use the terms *situations* and *contexts* interchangeably and use the term *domain* to refer to trait-relevant categories of situations.

---

[2] The constructs of self-control, self-regulation, and self-discipline, as well as of impulsiveness and impulsivity, are defined differently by different authors (Duckworth & Kern, 2011; Evenden, 1999; Olson, Schilling, & Bates, 1999). There is, however, considerable overlap. Most definitions of self-control and synonymous constructs refer to the ability to regulate impulses in the service of long-term goals and standards; most definitions of impulsiveness and impulsivity connote the obverse.

[3] Although we do not necessarily agree with Blasi's claim, we include this reference to emphasize that experts in the field recognize self-control as being central to moral character.

STUDIES IN THE NATURE OF CHARACTER: HARTSHORNE AND MAY, 1928–1930

In the late 1920s, Hartshorne and May conducted an impressive observational study of character with several thousand school children. The children were observed in a number of different situations to assess their moral character in terms of lying, cheating, stealing, selfishness versus altruism, and impulsivity versus self-control. Contrary to expectations, correlations between character measured in one situation and character measured in another situation were relatively low (around $r = .20–.30$). In other words, character appeared to be domain-specific. Students lied, cheated, and stole, and were selfish and impulsive in some situations but not others.

Foreshadowing our work, Hartshorne and May (1929) discussed domain-specific versus domain-general self-control, noting that,

> the chief difference . . . between the civilized and the uncivilized lies not in the possession of a general virtue of persistence or a general power of inhibition, but in the development of certain habits rather than others. The savage . . . is able to perform prodigious feats of endurance, whether of physical strain or of pain; he has acquired self-control in certain directions quite beyond the power of the civilized man who preaches to him of temperance and industry. In other respects, however, as in the accumulation of goods, or in the use of alcohol, he is less disciplined than those who make a virtue of property and sobriety. (p. 283)

Said differently, the "savage" might appear impulsive in the domains of alcohol and work, but self-controlled when it comes to physical endurance.

PERSONALITY AND ASSESSMENT: MISCHEL, 1968

Personality traits are often defined as stable patterns of behaving, thinking, and feeling that are relatively consistent across time and situation (McAdams & Pals, 2006). From a trait perspective, someone who is extroverted is assumed to be talkative and outgoing across domains, such as home, school, work, and play. In 1968, Walter Mischel contended in his book *Personality and Assessment* that psychologists had vastly overestimated the stability of personality across domains. Reviewing the extant literature on domain-generality in personality, Mischel (1968) found that the typical correlation between behavior in different domains was surprisingly low (about $r = .30$)[4] and concluded, "Although behavior patterns often may be stable, they usually are not highly

[4] A correlation of $r = .30$ indicates that 9% ($.30^2 = .09$) of the variance in one variable can be predicted by the other variable. In other words, behavior in one domain predicts 9% of the variance in behavior in another domain. Some people interpreted the finding that less than 10% of the variance overlapped to mean that personality did not exist, or, if it did, then it was not very important.

generalized across situations" (p. 282). *Personality and Assessment* incited what is now known as the *person-situation debate*. On one side, the *situationists* argued that so-called global or broad traits do not predict behavior very well and that situations are more important (Funder, 2004; Miller, 2014). Personality psychologists, on the other hand, maintained that personality was just as important as situations. Funder & Ozer (1983), for instance, demonstrated that the relationship between situations and behavior was similar in magnitude to the relationship between traits and behavior.

Personality psychologists also argued that personality is better at predicting overall (aggregate) patterns of behavior rather than specific instances of behavior in particular situations (Epstein & O'Brien, 1985). In other words, when we deem someone to be friendly, we regard them as being friendly across situations in general versus being friendly in specific contexts. While some personality psychologists have argued that traits are stable patterns of behavior over certain *types* of situations, rather than *all* situations (Bandura, 2006), in practice, most personality measures do not, in fact, specify types of contexts. The omission of context from most personality questionnaires requires the respondent to consider his overall level of behavior. One can imagine a respondent whose self-control is prodigious when it comes to finishing work assignments on time but minimal when it comes to kicking a smoking habit; faced with a question on his overall level of self-control, this individual might say "moderate." Mischel and Shoda (1995) have pointed out that such standard practices implicitly treat domain-specific deviations from mean levels of behavior as error or noise that obscures the "true" trait. By focusing on averages, this approach ignores potentially interesting domain-specific variance in behavior.

TRAITS AS DENSITY DISTRIBUTIONS OF STATES: FLEESON, 2001

In 2001, William Fleeson introduced a density distribution model of traits. From this perspective, traits can be thought of as distributions of states (i.e., the combined/overall set of states), where states are momentary actions, thoughts, and feelings. For instance, controlling one's temper, exercising, and eating in moderation are examples of states that reflect self-control, and all of an individual's self-control-related states together form a distribution. Each distribution has an average (i.e., typical or characteristic degree of the states) and variance (i.e., differences in the degree of the states: some are lower or higher than the average). As a resolution to the person-situation debate, Fleeson demonstrated that individuals' average levels of traits are highly stable ($rs > .90$; i.e., someone who is self-controlled on average one week is likely to be self-controlled on average the next week), even though marked within-individual variance exists (and single states demonstrate relatively low test-retest stability). In other

words, while an individual might exhibit prodigious self-control at one time and behave impulsively at a later time, his or her average degree of self-control is stable over time.

Domain-Specific Impulsive Behavior

We believe that the search for coherence in personality lies not in the negation of the situation but in its active study. In doing so—and specifically in examining patterns of behavior that are consistent across certain types of situations—we might "reconcile the variability of the individual's behavior across situations with our intuitive conviction that each individual is characterized by stable and distinctive qualities" (Mischel, 2009, p. 284). Specifically, people can be characterized by distinctive and stable patterns of situation-behavior relations. In this view, personality is to be found in the patterns of actions, thoughts, and feelings that are stable, not across all situations, but, rather, certain *types* of situations (e.g., people might be impulsive in situations they find tempting but self-controlled in non-tempting situations). This position—and the need to examine within-individual differences in behavior—has been advocated by leaders in personality psychology (Fleeson, 2007; Funder, 2009; Lucas & Donnellan, 2009; Mischel, 2004; Roberts, 2007).

In order to fully understand personality traits and character, we must know the situations that facilitate or undermine them. Fleeson and Noftle (2008) have cautioned that "studying each reaction to each situation as an isolated variable would lead to slow progress on identifying processes" (p. 1677). What is needed, in our view, is a taxonomy of situations that emerges from *psychologically meaningful*, as opposed to nominally convenient (e.g., work, love, play) domains (Zayas, Whitsett, Lee, Wilson, & Shoda, 2008). These domains (i.e., groupings of trait-relevant situations) may vary by trait; the organization of situations relevant to altruism (e.g., family, friends, strangers), for instance, may not be the same as that for self-control (e.g., food, sex, work). Once domains are identified, underlying characteristics can be identified and used to explain domain-specificity. Such an approach "analyzes situations into their psychologically relevant characteristics, treats situations as having degrees of a characteristic, allows situations to have multiple characteristics at once, studies characteristics independently, and produces situation-state contingencies with coefficients that have direction and graded magnitude" (Fleeson, 2007, p. 830). In other words, a given trait will have different trait-specific relevant situations distinguished by characteristics that are psychologically meaningful for the trait. For instance, relevant domains for extraversion might be home, school, work, and play. A psychologically meaningful characteristic of these situations might be "evokes (or challenges) self-confidence." Domain-specific

self-confidence might explain domain-specific extraversion: people are more extraverted in domains where they feel more confident.

What, then, explains domain-specificity in self-control? We considered two rival, but not mutually exclusive, possibilities. First, a person might be impulsive in one domain but not others because he is tempted in that domain but not others. For instance, Eliot Spitzer might be highly tempted when it comes to extramarital sex but not tempted at all when it comes to procrastinating or drinking. Alternatively, a person might be impulsive in one domain because he does not think it is bad. According to this explanation, Eliot Spitzer might not have thought that extramarital sex was bad. Consequently, we propose that subjective temptation and perceived harm[5] are psychologically meaningful factors related to self-control that can explain domain-specificity. Specifically, we suggest that both subjective temptation and perceived harm are domain specific, whereas self-control resources and strategies are domain general. Consistent with the empirical work by Baumeister and colleagues (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven & Baumeister, 2000), we assume that self-control processes rely on cognitive resources (e.g., attention, working memory) that are divisible, at least at any given point in time, and therefore judiciously allocated. As shown in figure 18.1a, we suggest that individuals who are more self-controlled across all domains relative to others have more self-control resources to deploy (e.g., greater working memory capacity) and more effective metacognitive strategies (e.g., precommitment, goal setting and planning, and psychological distancing). On the other hand, as illustrated in figure 18.1b, an individual's impulsive behavior will vary across domains as a function of his or her idiosyncratic, domain-specific, subjective evaluations of temptation and perceived harm. For example, Eliot Spitzer might have prodigious self-control resources and effective strategies for reducing the cost of resisting temptation. But, relative to other impulses that Spitzer did successfully resist (e.g., the urge to procrastinate), extramarital sex might have elicited exceptionally strong urges and/or been evaluated as relatively benign. In addition, we predicted that more-self-controlled individuals would be better at overcoming their idiosyncratic temptations and refraining from behaviors they deem harmful.

It is important to emphasize that we make a distinction between impulsive behavior and the psychological processes underlying the expression (or inhibition) of impulsive behavior (Duckworth & Steinberg, 2014). In our theoretic framework, impulsive behavior is behavior that generally brings about short-term gratification at the expense of longer-term goals; whereas,

[5] Although "bad" and "harmful" are not exactly the same, things that are bad are harmful (broadly speaking), and vice versa. In our investigation, we were primarily trying to tap a general negative/bad/avoid (vs. positive/good/approach) construct and the nuanced differences were not crucial in our theoretic model.

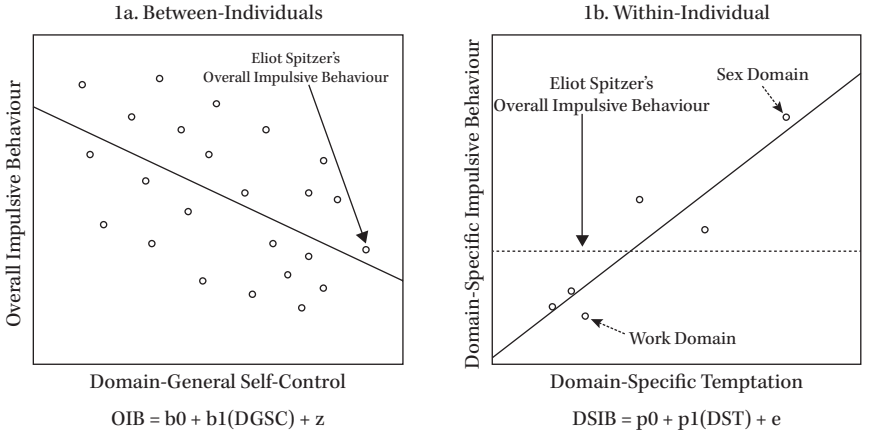

FIGURE 18.1 Stylistic illustration of a model explaining both (a) between-individual, domain-general variance and (b) within-individual, domain-specific variance in impulsive behavior.

OIB = overall Impulsive Behavior; DGSC = domain-general self-control; DSIB = domain-specific impulsive behavior; DST = domain-specific temptation.

temptation (i.e., impulse strength) is a psychological process that can affect the expression of impulsive behavior.[6]

DOMAIN-SPECIFIC IMPULSIVE BEHAVIOR IN ADULTS

In an initial set of studies (Tsukayama, et al., 2012), we tested our model in adults. The primary goals of these initial studies were to provide empirical evidence for domain-specificity and explain why impulsive behavior might be domain specific. In the first two studies, we developed and validated a self-report questionnaire assessing impulsive behavior in six domains: work, interpersonal relationships, drugs, food, exercise, and finances. We generated items through focus groups instructed to nominate behaviors that reflect impulsivity or self-control and, in addition, by examining existing measures of impulsivity and self-control (e.g., Tangney, Baumeister, & Boone, 2004; Whiteside & Lynam, 2001). Although we tried to include a broad sample, the items were not intended to be an exhaustive set of impulsive behaviors.

Exploratory and confirmatory factor analyses suggested that a domain-specific model fit the data better than a domain-general model.[7]

[6] Although "temptation" and "impulse strength" are not exactly the same, things that are tempting elicit strong impulses (broadly speaking), and vice versa. In our investigation, we were primarily trying to tap a visceral (vs. contemplative) approach construct and the nuanced differences were not crucial in our theoretic model.

[7] Factor analyses examine patterns of correlations among variables to indicate clusters of variables that covary together. If self-control were domain general, all of the items should covary with each other to about the same degree. If self-control were domain specific, then certain items (e.g., all of the food items) should covary more with each other than with non-related items (e.g., sex items).

The domain-specific subscales indicated by the factor analyses demonstrated convergent validity (i.e., correlated) with a widely used domain-general measure of self-control. In other words, each of the domain-specific impulsive behavior subscales were negatively correlated with domain-general self-control, indicating that they are actually measuring impulsive behavior. If self-control were domain general, then domain-specific measures should not predict above and beyond a general measure. However, the domain-specific subscales also predicted theoretically relevant outcomes above and beyond the domain-general measure. For instance, impulsive behavior in the food domain predicted body mass index over and beyond the domain-general self-control questionnaire. In addition, we found that the domain-specific variance (within-individual variance across domains) was greater than the domain-general (between-individual) variance. One way to think about this is that impulsive behavior is more domain specific than domain general. These first two studies provided empirical support for domain-specific impulsive behavior.

In a third study, we added measures of temptation (e.g., how tempted are you to procrastinate?) and perceived harm (e.g., how bad do you think it is to procrastinate?) in addition to the measures of domain-specific impulsive behavior (e.g., how often do you procrastinate?). This allowed us to test our hypothesized explanations for domain-specific impulsive behavior (i.e., domain-specific temptation and perceived harm). In support of our model, we found that individuals behaved more impulsively in domains that they found tempting and, to a much lesser extent, resisted vices that they perceived as harmful. Specifically, temptation and perceived harm together explained 59% of the within-individual variance across domains; whereas temptation explained 40% of the unique variance compared with 2% for perceived harm. On its own, temptation explained most of the variance (57%) in domain-specific impulsive behavior. Finally, in a fourth study, we tested the hypothesis that temptation is highly domain-specific (i.e., what is tempting to one individual can hold no appeal whatsoever to another) by recruiting participants in special interest groups (e.g., procrastinators) on Facebook. We confirmed that participants who were especially tempted in the target domain (e.g., work) were not likely to be more tempted in unrelated domains (e.g., food). These findings suggest that individuals are in "impulsive-problem" groups because of domain-specific temptation.

Altogether, these results provided preliminary support for our proposed model of impulsivity. Some individuals were more impulsive than others on average, and the tendency to be impulsive in one domain (e.g., finances) correlated with the tendency to be impulsive in other domains (e.g., food and drugs). But, the average individual also varied dramatically in how gratifying he or she found particular activities, and consequently, in the relative frequency with which he or she indulged in them. In contrast, the perceived

harm associated with indulging was a weaker predictor of impulsive behavior. Moreover, as predicted, temptation had a weaker effect and perceived harm had a stronger effect for individuals with higher self-control. That is, more self-controlled individuals are better at overcoming their idiosyncratic temptations and refraining from behaviors they deem harmful.

DOMAIN-SPECIFIC TEMPORAL DISCOUNTING

In another investigation (Tsukayama & Duckworth, 2010), we explored domain-specificity through the temporal discounting paradigm. Whereas self-report measures of impulsivity are sensitive to social-desirability biases, choices in temporal discounting (sooner-smaller versus later-larger rewards) are not as transparent. For instance, participants were asked to choose between $8 immediately and $16 in three years. Participants made hypothetical choices between sooner-smaller versus later-larger rewards of money, chips, candy, and beer. Corroborating our studies using self-report questionnaires, we found that temporal discounting was domain specific, and domain-specificity in temptation partially explained domain-specificity in temporal discounting.

New Findings on Domain-Specific Impulsivity

We see three major limitations of our initial work on domain-specificity in impulsivity. First, the use of non-random samples of college students constrained the generalizability of our results, especially regarding age, socioeconomic status, and ethnicity. Second, the cross-sectional data did not allow us to demonstrate temporal consistency (e.g., test-retest reliability/stability). Finally, the cross-sectional nature of the study limited causal inference (e.g., do subjective temptation and perceived harm cause, or just correlate with, impulsive acts?).

With a generous grant from the John Templeton Foundation and the Character Project at Wake Forest University, we were able to conduct a longitudinal study of domain-specific impulsivity with cohorts of varying ages (e.g., childhood, adolescence, early adulthood, and late adulthood) at baseline. This allowed us to (1) replicate and extend our initial findings across developmental periods with cohorts that were more socioeconomically and ethnically diverse than our initial sample; (2) establish test-retest stability (i.e., temporal consistency); (3) strengthen causal inference through a longitudinal design; and (4) explore gender differences.

In order for our measures to more quintessentially represent morality than our current domains, we added items to our domain-specific questionnaires (on behavior, temptation, and harm) so that the subscales reflected the seven deadly sins (wrath, envy, pride, greed, lust, sloth, and gluttony; see table 18.1). We also included informant-report measures of moral character to

TABLE 18.1 } **Domains from Initial Studies and New Domains Developed to Represent the Seven Deadly Sins**

Old Domains	New Domains
1. Work	1. Work (Sloth)
2. Exercise	2. Exercise (Sloth)
3. Food	3. Food (Gluttony)
4. Drugs	4. Drugs (Gluttony)
5. Finance	5. Finance (Gluttony)
6. Interpersonal	6. Wrath
6. Interpersonal	7. Envy
6. Interpersonal	8. Pride
6. Interpersonal	9. Greed
6. Interpersonal	10. Lust

Note. The original Domain-Specific Impulsivity Scale (DSIS) had six domains, whereas the revised scale (DSIS-7DS) has ten domains.

demonstrate that the self-report ratings represent more than just people's view of their own character.

Four cohorts (school-age children, adolescents, young adults, and older adults) completed the domain-specific impulsivity scale at four time points over two years, endorsing each item separately using three different prompts: frequency of behavior, perceived harm of behavior, and perceived temptation/enjoyment of behavior. (Children and adolescents, however, did not have questions pertaining to lust and drugs.) Participants at each time point also completed a domain-general self-control questionnaire (Tangney et al., 2004). Self-report questionnaires (for adults) and school record data (for children and adolescents) were used to assess age-appropriate, theoretically predicted outcomes (e.g., grade point average, physical health).

Our findings, described below, address the following questions related to domain-specific impulsive behavior. Is there evidence for domain-specific impulsive behavior? If so, what explains the domain-specificity? Do these findings generalize across the life-span? Are the hypothesized explanations causal or merely correlational? If impulsive behavior is not trait-like in terms of situational consistency, is it trait-like in terms of temporal consistency? Even if there are no gender differences in domain-general impulsive behavior, could there be domain-specific gender differences?

FACTOR STRUCTURE OF THE REVISED DOMAIN-SPECIFIC IMPULSIVITY SCALE

Confirmatory factor analyses verified the hypothesized factor structure for a revised domain-specific impulsive behavior scale (DSIS-7DS) that assesses the seven deadly sins (Tsukayama, Von Culin, & Duckworth, 2015b).

Specifically, ten domains were identified: work (sloth), exercise (sloth), food (gluttony), drugs and alcohol (gluttony), finance (gluttony), wrath, envy, lust, greed, and pride.[8] The hypothesized domain-specific factor structure fit the data better than a domain-general one-factor model.

CORRELATIONS WITH INFORMANT-REPORT RATINGS OF CHARACTER

Self-ratings of character were corroborated by informant ratings (Tsukayama et al., 2015b). Specifically, the average correlation between corresponding domain-specific impulsive-behavior subscales and informant-report ratings of the seven heavenly virtues (e.g., impulsive pride and the corresponding virtue of humility) were higher than their associations with non-corresponding virtues (e.g., impulsive pride and the non-corresponding virtue of chastity). These findings suggest that the participants' self-report ratings did not just reflect subjective and possibly inaccurate perceptions of themselves.

CAUSAL ROLE OF TEMPTATION IN IMPULSIVITY

While domain-specific temptation is strongly correlated with domain-specific impulsive behavior, analyses establishing that changes in temptation predict changes in behavior would provide a more rigorous test of causality (Tsukayama, Patrick, & Duckworth, 2015a).[9] As hypothesized, changes in domain-specific temptation robustly predicted changes in domain-specific impulsive behavior, whereas changes in perceived harm only predicted changes in behavior in about half of the analyses. In other words, if an individual's temptation to impulse shop increased, she or he was more likely to engage in impulse shopping. However, changes in the perceived harm of impulse shopping did not consistently make the likelihood of impulse shopping more or less likely. These results suggest that domain-specific temptation explains domain-specific impulsive behavior in a causal (and not just statistical) sense.

---

[8] The astute reader might notice that the first half of the list (work, exercise, food, drugs and alcohol, and finance) seems to be more related to the external world; whereas the second half (wrath, envy, lust, greed, and pride) seems more psychological/emotional. In the initial scale, "interpersonal/emotional" was a single domain. In the revised scale, we specifically created new items to expand the interpersonal/emotional domain to include wrath, envy, lust, greed, and pride. These new items formed their own separate domains (rather than a single interpersonal/emotional domain).

[9] This "fixed-effects" approach (Allison, 2009; Duckworth, Tsukayama, & May, 2010) controls for time-invariant third-variable confounds because a confounding variable would have to predict these changes and also be tightly yoked to changes in the predictor over time (i.e., the confound and predictor would have to change together in synchrony over time). If a third variable does not change over time, then changes in the outcome cannot be caused by the third variable. Thus, time-invariant, between-individual variables cannot confound the relationship. Basically, each person serves as his or her own control in a fixed-effects model.

DOMAIN-SPECIFIC IMPULSIVE BEHAVIOR MODEL GENERALIZES ACROSS THE LIFESPAN

We tested our model of impulsive behavior in separate analyses for each cohort and found that impulsive behavior is domain specific (Tsukayama et al., 2015b). Specifically, we found substantial domain-specific (within individuals) *and* domain-general (between-individual) variance in impulsive behavior, but the domain-specific variance was substantially larger than the domain-general variance in overall impulsive behavior. The domain-specificity (i.e., within-individual variance across domains) was explained by domain-specific subjective temptation and perceived harm; whereas domain-general impulsive behavior (between-individual variance) was explained by domain-general self-control. As predicted, more self-controlled individuals were less likely to engage in behaviors they found tempting and were more likely to avoid behaviors they perceived as harmful.

RANK-ORDER STABILITY AND STABILITY OF DOMAIN-SPECIFIC IMPULSIVITY PROFILES OVER TIME

While we found that impulsive behavior is not highly consistent across domains, we predicted that it would be consistent across time (Tsukayama & Duckworth, 2015b). Both domain-general self-control ability and domain-specific impulsive behavior demonstrated relatively high six-month test-retest stability. In addition, we conducted profiles analyses (Furr, 2008) to examine whether the within-individual pattern of impulsive behavior across domains was stable across time. Profile analysis examines the correlation between patterns. In this case, we examined the pattern/profile of domain-specific impulsive behavior for each individual at one time point and compared it to the profile at a later time point (see figure 18.2). For example, Young Adult number 117 in is relatively impulsive in certain domains (e.g., work and exercise) but not others (e.g., wrath and drugs), and this pattern (i.e., relative rank ordering) is stable across time. We found high stability for the average overall agreement profiles. These results suggest that while impulsive behavior might not be trait-like in terms of consistency across domains, impulsive behavior is trait-like in terms of consistent patterns across time.figure 18.2

GENDER DIFFERENCES IN DOMAIN-SPECIFIC IMPULSIVE BEHAVIOR

A domain-general perspective is constrained to examining gender differences in overall impulsive behavior (Tsukayama & Duckworth, 2015a), and extant evidence suggests that men and women do not differ (Feingold, 1994). However, the question of whether men or women are more impulsive may depend

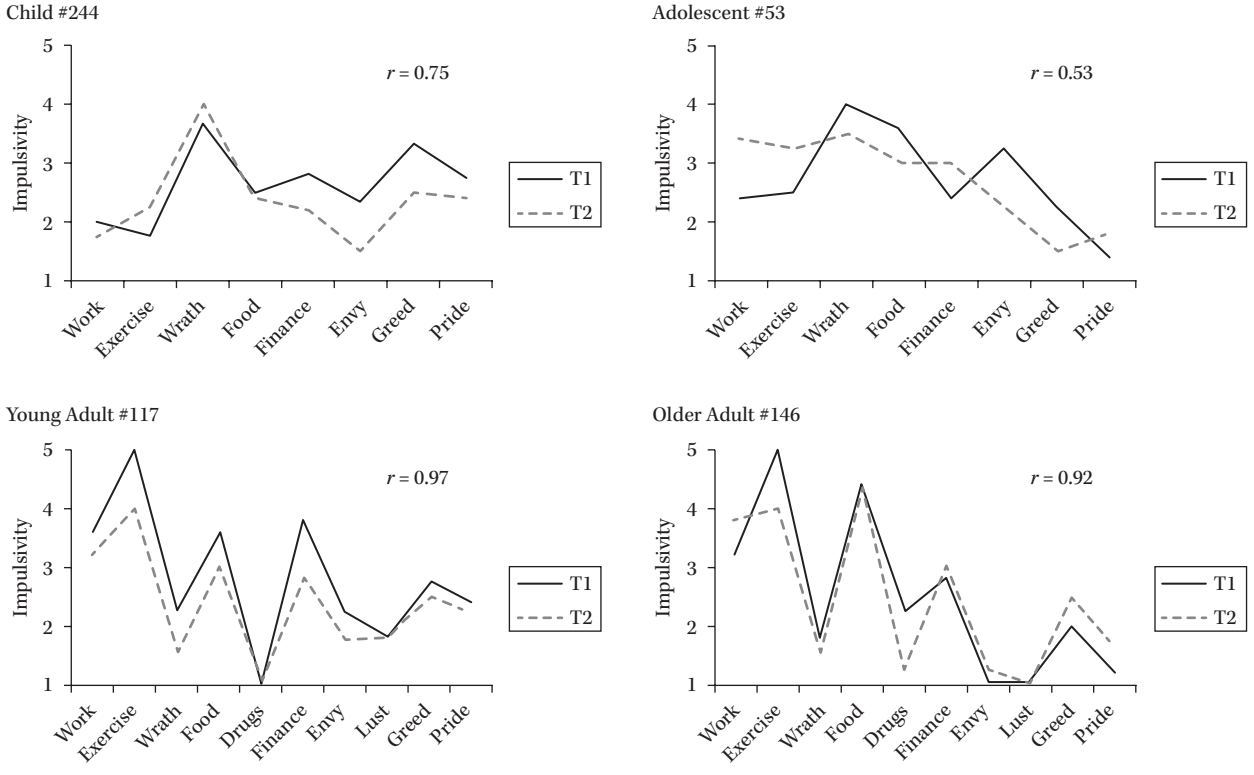

FIGURE 18.2. Overall agreement profiles illustrating the high stability of domain-specific impulsive behavior profiles across a six-month period. One participant was randomly selected from each cohort.

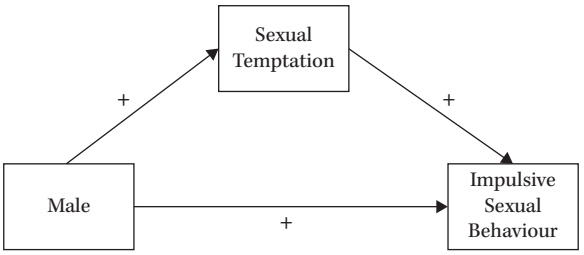

FIGURE 18.3 Illustration of a mediation model, where men's tendency to engage in more impulsive sexual behavior compared to women is explained by men's greater temptation to engage in sexual behavior.

on the type of behavior in question. In other words, there might be gender differences in specific domains. For instance, men might be more impulsive in the sex domain; whereas women might be more impulsive in the shopping (or finance) domain.

In the different age groups—children, adolescents, young adults, and older adults—males and females rated themselves comparably in domain-general self-control ability. Similarly, there were no gender differences in overall impulsive behavior, except adolescent females reported more impulsive behavior compared to adolescent males. As predicted, we found theoretically sensible gender differences in domain-specific behavior, and these differences were mediated by corresponding differences in domain-specific temptation. For instance, men were more likely to engage in impulsive sexual behavior, and this behavior was mediated by higher levels of temptation (see figure 18.3). This suggests that men are more impulsive than women are when it comes to sex because men are more tempted in the sexual domain

Conclusions and Future Directions

Is impulsivity domain-specific? The studies summarized here suggest that it is. In particular, these studies collectively (1) establish a model that explains both domain-general and domain-specific impulsive behavior that is generalizable across development periods, (2) demonstrate temporal consistency of domain-general self-control and domain-specific impulsive behavior, and (3) provide an explanation for gender differences in impulsive behavior.

More broadly, the reviewed research sheds light on putative inconsistencies in moral character. Because the domains in the more recent investigations reflect the seven deadly sins, this investigation has direct relevance to moral behavior. Although our scale does not include every potential moral infraction, our investigation has broad implications for moral psychology because self-control underlies moral character in general (Baumeister, 2005). For

instance, self-control is relevant to all five moral foundations of Haidt's (Haidt & Graham, 2007) taxonomy: care for others (e.g., overcoming selfish desires), justice and fairness (e.g., resisting the temptation to cheat), in-group loyalty (e.g., making self-sacrifices for the good of the group), respect for authority/tradition (e.g., submitting to authority), and purity/sanctity (e.g., avoiding bestiality).

Future directions include discovering the developmental origins of desires (i.e., temptation) and perceived harm as well as developing interventions to manipulate temptation and perceived harm in order to further strengthen causal inference and potentially promote moral character. One of the major limitations of the reviewed studies is the reliance on self-report measures. Future studies should explore alternatives to self-report questionnaires to assess temptation, perceived harm, and impulsive behavior.

We agree with Epstein and O'Brien (1985) that "behavior is unquestionably to some extent general and to some extent specific, and one can choose to study one aspect or the other" (p. 513). Unfortunately, it seems that personality research has in general focused on the former. As summarized in this review, examining domain-specific aspects of behavior revealed that the influence of context on behavior is both substantial and systematic. Although we place special emphasis on domain-specificity, we do not wish to downplay the importance of domain-general processes. Both aspects are important. We accentuate the domain-specific aspect because it has largely been ignored. Although interest in domain-specificity has been increasing (e.g., Goldstein & Weber, 1997; Shoda, Mischel, & Wright, 1994; Weber, Blais, & Betz, 2002), empirical domain-specific studies are still relatively rare. In order to obtain a more complete understanding of character, domain-specificity must be accounted for instead of ignored. Collectively, the findings summarized in this review highlight the utility of a domain-specific approach— namely, in terms of improved prediction of life outcomes, understanding psychological processes, and targeted interventions.

Funder (2009) has pointed out that a focus on within-person variance across domains raises two questions: "Where do these patterns come from? How are they important?" (p. 122). Likewise, Mischel and colleagues have asserted, "By addressing not only the average level of behavior (e.g., overall agreeableness) but also when, where, and with whom it occurs, one can see the individual's distinctive coherent and systematic patterns of behavior variation and glimpse the psychological processes and person variables that underlie them" (Shoda, et al., 1994, p. 686). We agree. Whereas the standard practice of averaging across domains can obscure important individual differences (Mischel, Shoda, & Mendoza-Denton, 2002), examining domains as variables of interest can provide a more nuanced and accurate view of personality. As the studies summarized in this review have demonstrated, psychologically salient aspects of the domain (temptation, in this investigation, and perceived harm, to a lesser degree) can, to a large extent,

explain variance in domain-specific behavior. Furthermore, we suggest that simultaneously examining both within-individual and between-individual variance can lead to important insights. For example, the self-control by temptation and self-control by perceived harm cross-level interactions revealed that more self-controlled individuals are better at overcoming temptations they found alluring yet harmful.

These studies provide an explanation for the variation in impulsive behavior within individuals across domains: domain-specific impulsive behavior is a result of domain-specific temptation and, to a lesser extent, perceived harm. While it is possible that domain-specific impulse-control systems or the opportunity to engage in impulsive behavior explains domain-specificity, these studies suggests that the lack of domain-general self-control combined with domain-specific temptation (and other potential factors) can give rise to domain-specific impulsive behavior. Consistent with this explanation, a recent review of neuroimaging studies suggests that prefrontal brain regions supporting self-control are domain general, whereas the subcortical regions representing impulses (i.e., "the reward, salience, and emotional value of a stimulus") vary depending on the domain (Heatherton & Wagner, 2011, p. 134). These results are congruent with Mischel's (1973) proposal that an individual's "inconsistent" behavior across domains is a function of the individual's construal of the situation. If an individual subjectively perceives temptation to be high (or harm to be low), then that individual is more likely to engage in impulsive behavior.

So, why was Eliot Spitzer generally self-controlled, yet impulsive when it came to extramarital sex? Our studies suggest that he had stronger impulses to engage in impulsive sexual behavior than to engage in other impulsive behaviors.[10]

References

Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles, CA: Sage.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy Beliefs of Adolescents* (pp. 307–337). Greenwich, CT: Information Age Publishing.

Baumeister, R. F. (2005). Self and volition. In W. R. Miller & H. D. Delaney (Eds.), *Judeo-Christian perspectives on psychology* (pp. 57–72). Washington, DC: American Psychological Association.

[10] The authors contributed equally to this work. This project/publication was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation, and Grant K01-AG033182 from the National Institute on Aging. We also thank Christian Miller, Mike Furr, and Will Fleeson for their helpful comments on earlier drafts of this chapter, as well as the audiences at the 2012 and 2013 Character Project conferences for their insightful comments on this project. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.

Duckworth, A. L., & Kern, M. (2011). Meta-analytic evidence for the convergent validity of self-control measures. *Journal of Research in Personality*, *45*, 259–268.

Duckworth, A.L., & Steinberg, L. (2015). Unpacking self-control. *Child Development Perspectives, 9*(1), 32–37.

Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychology and Personality Science*, *1*, 311–317.

Epstein, S., & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin*, *98*, 513–537.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology. Special Issue: Impulsivity*, *146*, 348–361.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*, 429–456.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027.

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, *75*, 825–862.

Fleeson, W., & Noftle, E. E. (2008). The end of the person-situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*, 1667–1684.

Funder, D. C. (2004). *The personality puzzle* (3rd ed.). New York, NY: Norton.

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality*, *43*, 120–126.

Funder, D. C., & Ozer, D. J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology*, *44*, 107–112.

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*, 1267–1316.

Goldstein, W. M., & Weber, E. U. (1997). Content and discontent: Indications and implications of domain specificity in preferential decision making. In W. M. Goldstein & R. Hogarth (Eds.), *Research in Judgement and Decision Making* (pp. 566–617). Cambridge: Cambridge University Press.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98–116.

Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Science*, *15*, 132–139.

Lapsley, D. K., & Narvaez, D. (2006). Character education. In A. Renninger & I. Siegel (Eds.), *Handbook of child psychology: Child Psychology in Practice* (6th ed., Vol. 4, pp. 248–296). Hoboken, NJ: John Wiley & Sons.

Lucas, R. E., & Donnellan, M. B. (2009). If the person-situation debate is really over, why does it still generate so much negative affect? *Journal of Research in Personality*, *43*, 146–149.

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: fundamental principles for an integrative science of personality. *The American Psychologist*, *61*, 204–217.

Miller, C. B. (2014). *Character and Moral Psychology*. Oxford: Oxford University Press.

Mischel, W. (1968). *Personality and assessment*. Hoboken, NJ: John Wiley & Sons.

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*, 252–283.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1–22.

Mischel, W. (2009). From personality and assessment (1968) to personality science, 2009. *Journal of Research in Personality*, *43*, 282–290.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268.

Mischel, W., Shoda, Y., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science*, *11*, 50–54.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247–259.

Olson, S. L., Schilling, E. M., & Bates, J. E. (1999). Measurement of impulsivity: Construct coherence, longitudinal stability, and relationship with externalizing problems in middle childhood and adolescence. *Journal of Abnormal Child Psychology*, *27*, 151–165.

Roberts, B. W. (2007). Contextualizing personality psychology. *Journal of Personality*, *75*, 1071–1082.

Shoda, Y., Mischel, W., & Wright, J. C. (1994). lntra-individual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, *67*, 674–687.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*, 271–322.

Tsukayama, E., & Duckworth, A. L. (2010). Domain-specific temporal discounting and temptation. *Judgment and Decision Making*, *5*, 72–82.

Tsukayama, E., & Duckworth, A. L. (2015a). *Gender differences in domain-specific impulsivity*. Manuscript in preparation.

Tsukayama, E., & Duckworth, A. L. (2015b). *Rank-order and ipsative profile stability of domain-specific impulsivity over time*. Manuscript in preparation.

Tsukayama, E., Duckworth, A. L., & Kim, B. (2012). Resisting everything except temptation: Evidence and an explanation for domain-specific impulsivity. *European Journal of Personality*, *26*, 318–334.

Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-specific impulsivity in school-age children. *Developmental Science*, *16*, 879–893.

Tsukayama, E., Patrick, S. D., & Duckworth, A. L. (2015a). *Lead us not into temptation: Changes in temptation predict changes in impulsive behavior*. Manuscript in preparation.

Tsukayama, E., Von Culin, K., & Duckworth, A. L. (2015b). *Domain-specific impulsivity across the lifespan.* Manuscript in preparation.

Weber, E. U., Blais, A.-R. e., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*, 263–290.

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*, 669–689.

Zayas, V., Whitsett, D. D., Lee, J. J. Y., Wilson, N., & Shoda, Y. (2008). From situation assessment to personality: Building a social cognitive model of personality. In G. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *Handbook of personality theory and testing* (Vol. 2), 337–402. Newbury Park: SAGE Publications.

19 }

Can Text Messages Make People Kinder?

Sara Konrath

This chapter asks some age-old questions that are bigger than the specific one we address in our research: can old dogs learn new tricks, can leopards change their spots, and do old habits die hard? That is, is it possible to change something as ingrained as someone's character? I specifically examine whether it is possible to change *empathic* character traits, using a rather unorthodox tool: the mobile phone.

There are reasons why the mobile phone might be an unlikely candidate for changing empathy. Mobile phones are often used to strategically block out in-person social interactions. During unwanted or awkward face-to-face conversations, people use them to escape psychologically, and they can also be used to avoid such interactions in the first place by screening or blocking calls. However, they also serve as potentially powerful social connectors. They can help to lubricate existing social relationships by keeping people in touch via regular check-ins. One other use of mobile phones that may not immediately come to mind is as habit formers. Given the ubiquity of mobile phones and their proximity to users, they have also successfully been used as powerful tools to inject people with regular pieces of useful information (e.g. health and safety messages) and thereby change their behaviors.

But can they be used to shape a person's character? This chapter will give some initial answers to this question, and it will also demonstrate that the answer depends on how we measure and define character. Although the evidence is not crystal clear or simple, this is the first known study that uses mobile technology to try to shift deep character traits in people, and I hope it won't be the last. This new use of a familiar technology offers a powerful tool to better understand the nature of character and the possibility of character change.

Defining Empathic Character

Empathy is very difficult to define and measure, since scholars apply the term loosely to refer to traits (Davis, 1983c), emotional responses (Batson, 2011), cognitive states or abilities (Davis & Kraus, 1997; Goldstein & Winner, 2012; Ickes, 1997; Van Honk et al., 2011), hypothetical responses (Barkai & Fine, 1983; Truax et al., 1966), and sometimes even behaviors (Lobchuk & Bokhari, 2008; Locraft & Teglasi, 1997; Strayer, 1980; Teherani, Hauer, & O'Sullivan, 2008). Given this broad usage in the literature, I define *empathic character* as encompassing interrelated dispositions, beliefs, skills, motivations, emotional responses, and behaviors. To some extent, this is in line with recent theorizing on the complex nature of character traits. For example, in his recent book on moral character, Christian Miller (2013) defines a character trait as consisting "of some cluster of . . . interrelated mental state dispositions" (p. 10). I see behaviors as important potential manifestations of such "interrelated mental state dispositions" of character. Although it is difficult to determine one's inner character from a simple observation of behavior, in combination with measures of more internal states, behaviors add an important dimension. For example, if a person felt a lot of compassion and also saw themselves as empathic, I find it doubtful to believe that he would be seen as possessing an empathic character unless he also regularly helped others and shared his time and belongings with them. Character is as character does.

Empathic character could also be called "kindness"; however, the term *empathy* is more commonly used within the psychological literature. For example, a keyword search on empathy versus kindness in the American Psychological Association's PsycInfo database (www.apa.org/psycinfo/) reveals that empathy is used 17 times more often than kindness (17,698 articles vs. 1,030 articles). People with empathic characters habitually feel concern and compassion for others, limit their own personal distress and other barriers to becoming involved in others' situations, and consistently try to do what they can to help and support others, regardless of who sees them or whether there are rewards to gain. They give and help and share because they want to help others in need, rather than to feel good or to receive something in return (Batson, 2011; Miller, 2013, chap. 5).

The Importance of Empathic Character

Empathic character traits are of central importance to people. They are readily identified in others (Wojciszke, 2005), and different people can agree on who scores higher versus lower in them (Bar-Tal & Raviv, 1979; Hartshorne, May, & Maller, 1928–1930). Indeed, both men and women say that the number one

attribute they look for in a romantic partner is kindness, even though there are gender differences lower in their preference lists when it comes to other attributes such as physical attractiveness (Buss & Barnes, 1986; Sprecher & Regan, 2002).

EMPATHIC CHARACTER IS ASSOCIATED WITH MORE PROSOCIAL BEHAVIOR

Numerous studies have found that people who score high on measures of empathic character traits tend to self-report and also exhibit more everyday prosocial behaviors; (Davis, 1983b; Kerber, 1984; Krebs & Sturrup, 1982; Larrieu & Mussen, 1986; McNeely & Meglino, 1994; Mehrabian & Epstein, 1972; Penner & Orom, 2009; Rushton, 1980, 1981; Rushton, Brainerd, & Pressley, 1983; Rushton, Chrisjohn, & Fekken, 1981; Rushton, Fulker, Neale, Nias, & Eysenck, 1989; Staub, 1974). Of course, there are also some null results in the literature (Amato, 1990; Yakimovich & Saltz, 1971), and probably many more in researchers' file drawers. However, meta-analyses find that there are overall positive relationships between empathic traits and prosocial behaviors (Borman, Penner, Allen, & Motowidlo, 2001; Eisenberg & Miller, 1987; Underwood & Moore, 1982).

Empathic traits are also correlated with longer-term, more deliberate prosocial behaviors, such as volunteering for non-profit organizations (Carlo, Okun, Knight, & de Guzman, 2005; Penner, 2002; Penner, Fritzsche, Craiger, & Freifeld, 1995; Stolinski, Ryan, Hausmann, & Wernli, 2004; Unger & Thumuluri, 1997). Even when controlling for a number of potential explanatory variables, empathic character traits are associated with increased volunteering behavior, charitable donations, and other prosocial behaviors in nationally representative Dutch and American samples (Bekkers, 2005, 2006; Clary, Snyder, & Stukas, 1996; Wilhelm & Bekkers, 2010).

There has also been some research examining empathic traits in more drastic situations. Indeed, Holocaust rescuers score higher in trait measures of empathy than non-rescuers and are more likely to be involved in prosocial behaviors several decades later (Fagin-Jones & Midlarsky, 2007; Midlarsky, Fagin Jones, & Corley, 2005; Oliner & Oliner, 1992). Other research finds that people who were awarded a national medal for exemplary caring score higher on empathic character traits than those who received a medal for bravery (Walker & Frimer, 2007). However, other studies examining people who helped in extreme situations (e.g., during car accidents or crimes) find mixed (Bierhoff, Klein, & Kramp, 1991) or null results (Huston, Ruggiero, Conner, & Geis, 1981). Meta-analyses would be useful in this literature.

Still, taken together, there is evidence that people with empathic character traits have an elevated probability of acting in the interests of others across a variety of situations.

EMPATHIC CHARACTER IS ASSOCIATED WITH OPTIMAL WELL-BEING AND HEALTH

It is well understood that empathic character predicts prosocial behavior, but what may be less known is that it has implications for the psychological well-being, mental health, and physical health of both the givers and receivers of care (Konrath & Brown, 2012).

More empathic or compassionate people experience less stress, anxiety, burnout, and depression (Au, Wong, Lai, & Chan, 2011; Burtson & Stichler, 2010; Dyrbye et al., 2010; Ironson et al., 2002; Steffen & Masters, 2005; Wink & Dillon, 2002). Empathic people have a lower physiological stress response to social evaluation (Ho, Konrath, Brown, & Swain, 2014), and this stress-buffering effect is present even when people are induced to be in temporary empathic states (Abelson et al., 2014; McClelland & Krishnit, 1988), suggesting that empathy causally affects stress regulation. People with more empathic character traits are also prone to healthier lifestyles, such as less alcohol and tobacco consumption (Adams, 2010; Kalliopuska, 1992), and not surprisingly, they also have better physical health (Dillon & Wink, 2007; Ironson, 2007; Konrath & Fuhrel-Forbis, 2011b). In addition, high empathy in doctors, therapists, and teachers is associated with better outcomes in their patients and students (Aspy & Roebuck, 1972; Beck, Daughtridge, & Sloane, 2002; Chang, Berger, & Chang, 1981; Coffman, 1981; Derksen, Bensing, & Lagro-Janssen, 2013; Kurtz & Grummon, 1972; Truax, et al., 1966; Waxman, 1983; West et al., 2006).

Is Empathy Changeable?

There is evidence that empathic character is both a stable disposition and also responsive to people's general environments and immediate contexts.

THE STABILITY OF EMPATHY

Heritability. Individual differences in empathy-related traits are at least partially heritable (Rushton, Fulker, Neale, Nias, & Eysenck, 1986), especially when considering their more affective aspects such as empathic concern (Davis, Luce, & Kraus, 1994; Matthews, Batson, Horn, & Rosenman, 1981; Zahn-Waxler, Robinson, & Emde, 1992).

Stability over time and situation. People who have empathic traits or behaviors at one time point, are also more likely to have empathic traits or behaviors at a later time point (Bar-Tal & Raviv, 1979; Davis, 1980; Davis & Franzoi, 1991; Eisenberg, Carlo, Murphy, & Court, 1995; Eisenberg et al, 2002; Eisenberg et al., 1987; Grühn, Rebucal, Diehl, Lumley, & Labouvie-Vief, 2008;

Penner, et al., 1995; Van der Mark, Ijzendoorn, & Bakermans-Kranenburg, 2002; Volbrecht, Lemery-Chalfant, Aksan, Zahn-Waxler, & Goldsmith, 2007). For example, one study found that preschoolers who spontaneously shared with other children were more likely to (i) behave prosocially, (ii) be rated by others as high in empathy, and (iii) see themselves as more empathic and prosocial in later childhood, adolescence, and young adulthood (Eisenberg et al., 1999). The authors conclude that "there is a prosocial personality disposition that emerges early and is somewhat consistent over time" (p. 1368). Next, in terms of cross-target consistency, children who show empathy to their mothers are also more likely to show empathy toward strangers (Robinson, Zahn-Waxler, & Emde, 2001; Young, Fox, & Zahn-Waxler, 1999).

THE MALLEABILITY OF EMPATHY

Situational factors affecting empathy. Yet there is also much research finding that empathy is malleable. For example, situational factors can affect the probability of prosocial behaviors (Baron, 1997; Darley & Batson, 1973; Isen, Horn, & Rosenhan, 1973; Isen & Levin, 1972; Krebs & Miller, 1985; Latané & Nida, 1981; Staub, 1974; Twenge, Baumeister, DeWall, Ciarocco, & Bartels, 2007). Since one sign of empathic character is consistent prosocial behavior, it is important to realize how much these types of behaviors are influenced by variables outside the person.

In addition, one expression of empathic character is the ability to identify others' emotions. Research has found that this can be affected by motivation. When people are more motivated because there is a chance to gain a financial reward, they can more accurately infer others' emotional states (Ickes & Simpson, 2003; Klein & Hodges, 2001). Moreover, temporal trends in dispositional empathy over the past 30 years suggest that it can be shaped by broad sociocultural factors (Konrath, O'Brien, & Hsing, 2011).

Empathic character can be directly taught and learned from others. Parents who are warmer, use more other-oriented discipline strategies (e.g., "Imagine how he must feel"), and discourage aggressive behavior elicit higher dispositional empathy and more prosocial behaviors from their children (Koestner, Franz, & Weinberger, 1990; Krevans & Gibbs, 1996; Strayer & Roberts, 2004; Zhou et al., 2002). Socio-emotional intervention programs can also increase empathic tendencies. For example, empathy can increase when similarity between the self and others is emphasized (Barnett, 1984; Brehm, Fletcher, & West, 1981), after role-playing another person's feelings or situation (Barak, Engle, Katzir, & Fisher, 1987; Feshbach, 1984; Feshbach & Konrad, 2001; Underwood & Moore, 1982), after observing others' misfortunes (Barnett, Howard, Melton, & Dino, 1982; Perry, 1975), and after exposure to highly empathic role models (Bryan & Walbek, 1970; Kohn, 1991; Kremer & Dietzen, 1991; Perry, 1975).

New Technology as Both a Cause and Solution for Declining Empathy

Over the past several decades, there have been documented declines in face-to-face social interactions, and in young people's self-perceived ability to care and connect with others (Konrath, Chopik, Hsing, & O'Brien, 2014; Konrath et al., 2011; O'Brien, Konrath, Grühn, & Hagen, 2013; Putnam, 2000; Twenge & Foster, 2008; Twenge, Konrath, Foster, Campbell, & Bushman, 2008). For example, our research has found that the personality trait narcissism has risen since the early 1980s among American college students (Twenge & Foster, 2008; Twenge, et al., 2008), while dispositional empathic concern and perspective taking have simultaneously been declining (Konrath, et al., 2011). In addition, American college students are now more likely to report insecure attachment styles, saying, for example, that they are comfortable without close relationships (Konrath, et al., 2014).

Some scholars have suggested that new media and technologies at least partially account for such changes (Konrath, 2012; Turkle, 2012). This is a plausible hypothesis considering that narcissism is associated with more frequent social media use (Carpenter, 2012; Panek, Nardis, & Konrath, 2013), and that studies have found that mobile phones can reduce people's empathy, feelings of closeness, feelings of trust, and prosocial behaviors (Abraham, Pocheptsova, & Ferraro, 2012; Przybylski & Weinstein, 2013). Yet, even though mobile phones may distract us from face-to-face interactions under some circumstances, they also literally connect us to others, making us feel closer to loved ones when there is physical distance (Coyne, Stockdale, Busby, Iverson, & Grant, 2011; Faulkner & Culwin, 2005; Levitas, 2013; Padilla-Walker, Coyne, & Fraser, 2012; Pettigrew, 2009). Technologies are simply tools, and with the introduction of new technologies, there are often struggles to make sense of them and optimize their potential (Postman, 1992).

Public health researchers have realized the potential of mobile technologies for several years (Konrath, 2015). For example, they have found that sending text messages to people encouraging them to quit smoking, exercise more often, or better manage their blood glucose can improve physical health outcomes (Cole-Lewis & Kershaw, 2010; Fiordelli, Diviani, & Schulz, 2013; Fjeldsoe, Marshall, & Miller, 2009; Herbert, Owen, Pascarella, & Streisand, 2013; Krishna, Boren, & Balas, 2009; Liang et al., 2011; Militello, Kelly, & Melnyk, 2012; Park, Howie-Esquivel, & Dracup, 2014; Shaw & Bosworth, 2012; Whittaker et al., 2009). Text messages are ideal intervention tools. They are inexpensive, embedded within participants' everyday lives, and have broad dissemination power. Approximately 9/10 Americans own a mobile phone (Pew Research Center, 2012), and 6/10 of those are smartphones (Nielsen, 2013). One study found that people had their phones in the room with them 90% of the time, and within arm's reach 50% of the time (Dey et al., 2011).

Yet so far, text messaging interventions have been limited to physical health outcomes or serious mental health problems; there is virtually no research using text messages to change psychological traits or social behaviors (Konrath, 2015). With respect to empathy specifically, most prior empathy-induction studies rely on classroom or laboratory settings (Barak, et al., 1987; Barnett, 1984; Barnett, et al., 1982; Brehm, et al., 1981; Bryan & Walbek, 1970; Feshbach, 1984; Feshbach & Konrad, 2001). Yet some innovative research has found that empathy can be taught using videos (Kremer & Dietzen, 1991; Warner, 1984), demonstrating the potential effectiveness of more remote interventions.

The Current Study: Can Text Messages Make People Kinder?

The current study is the first known study that attempts to change character traits via text messaging. Our rationale was that daily practice may help to create empathic habits. Some theorists consider more automatic (effortless) versus more controlled (effortful) aspects of empathy (Hodges & Wegner, 1997). We posit that if people repeatedly engage in practicing empathy-related tasks, this could lead to more automatic (or habitual) empathic responses. This is in line with research demonstrating that tasks that initially take much concentration can later become automatic, almost done without thinking (Ericsson, 2006; Milton, Small, & Solodkin, 2004). Indeed, higher-empathy people exhibit spontaneous muscle mimicry of facial expressions when stimuli are presented below their conscious awareness (Sonnby-Borgström, Jönsson, & Svensson, 2003). This provides some support for the idea that some aspects of empathy may become more automatic with practice.

In this study, we randomly assigned participants to an empathy-building text-message group, an active control group or a passive (no intervention message) control group. We measured their empathic character traits before and after the intervention and also approximately six months later to examine whether the intervention had staying power. We found some evidence that this method can be successful. However, some of the results were unexpected and warrant further explanation. This study can help us to better understand the nature of empathic character and the changeability of different aspects of it.

Study Design and Measures

DEVELOPING THE TEXT MESSAGES

We developed the text messages by first researching the empathy literature for content ideas. Three commonly used components of empathic character arose from this process: (1) *Emotional empathy*, also known as empathic concern or compassion (Davis, 1983c; Galinsky, Maddux, Gilin, & White, 2008; O'Brien,

et al., 2013); (2) *cognitive empathy*, also known as perspective taking or role taking (Batson, Early, & Salvarani, 1997; Davis, 1983c; Galinsky, et al., 2008; O'Brien, et al., 2013); and (3) *prosocial behavior* as a potential consequence of empathy, such as helping or sharing (Eisenberg & Miller, 1987; Eisenberg et al., 1989). Control group messages were designed to capture both the emotional and the cognitive aspects of low empathy. These included messages that promoted loving oneself (i.e., high self-esteem) and messages that promoted psychological detachment or objectivity (i.e., psychological distance). Ideally, we would separate these two ways of being low in empathy into separate control groups; however, because of limited resources, they were combined into a single control group. We initially developed 39 empathy text messages and 38 self/objective control text messages.

Next, a group of 22 raters affiliated with our research lab scored each message on the extent to which it was other-oriented versus self-oriented, logical versus emotional, and behavioral versus in someone's mind. The final 21 empathy messages and 21 self/objective control messages were selected based on these ratings. An example empathy message was "Think about someone close to you. How can you show this person that you accept them and care for them?"; and an example self/objective message was "Think about your last social interaction. Did this person give you the level of respect you deserve?" Within each experimental condition, the messages were categorized as follows: eight emotional messages, eight cognitive messages, and five behavioral messages (see table 19.1). In the study, participants were sent these messages

TABLE 19.1 } **Example text messages**

Empathy Condition	**Self / Objective (Control) Condition**	
Think about somebody close to you. Do a small nice thing for this person today.	Think about all that you deserve. Do something nice for yourself today.	**Behavioral**
Smile at the next person you see, no matter who they are.	Smile at the mirror or window next time you see your reflection.	
Think about your last social interaction. What obstacles or challenges does the person face? See these problems from their point of view.	Think about your last social interaction. Think of two specific ways you could be more neutral and objective in the future.	**Cognitive**
Think about your last social interaction. What was your interaction partner really trying to say? Did you accurately read their body language?	Think about your last social interaction. What were you really trying to say. Did the other person accurately read your body language?	
Think about someone close to you. How can you show this person that you accept them and care for them?	Think about your last social interaction. Did this person give you the level of respect you deserve?	**Emotional**
Focus on connecting with and feeling close to the person you last interacted with, no matter who they are.	Focus on the ways that you are different and unique from the person you last interacted with, no matter who they are.	

© *Sara Konrath, 2013*

six times per day for 14 days, presented in a randomized order (i.e., four times per message).

Across both experimental conditions, the messages did not significantly differ in length. We also tried to match the messages in reading difficulty and were partially successful.

CHOOSING THE OUTCOME MEASURES

One of the problems we immediately encountered when planning this study was in determining how we could demonstrate that the text-message intervention had indeed increased empathic character. Given the numerous definitions and operationalizations of empathy in the literature, how could we most convincingly demonstrate whether our intervention worked? Would it be the most convincing if we found that people who received empathy-building text messages:

- experienced more feelings of compassion for someone in distress? (Batson, 2011);
- had more prosocial motives for helping others? (Konrath & Fuhrel-Forbis, 2011a);
- scored higher in self-reported dispositional measures of empathy, such as empathic concern and perspective taking? (Davis, 1983c);
- had stronger moral beliefs that it was important to help others? (Wilhelm & Bekkers, 2010);
- had stronger beliefs that it was unacceptable to behave aggressively? (Huesmann & Guerra, 1997);
- actually behaved more prosocially (less aggressively) in response to a provocation? (Konrath, Bushman, & Campbell, 2006; Taylor, 1967);
- were rated as more empathic by observers?;
- responded in prosocial ways to hypothetical scenarios or real situations in which others were in distress? (Batson, 2011; Truax et al., 1966);
- cooperated with other participants in a prisoner's dilemma game involving real money? (Balliet, 2010; Batson et al., 1995; Batson & Moran, 1999); or,
- gave more social support than they received? (Konrath, 2014).

Each of these different outcomes could be seen as acceptable evidence for the effectiveness of our intervention. Indeed, many intervention studies that have attempted to increase empathy or kindness only include one or two dependent measures. We thought it was better to use a constellation approach that went beyond demonstrating *whether* it was possible to alter empathic character using text messages. Instead, using this approach allowed us to examine *which*

specific aspects of empathic character or kindness were most responsive to the program. So we ended up including a number of different measures that other researchers have considered to be related to empathy or kindness.

STUDY DESIGN AND PARTICIPANTS

We used a pre- and post-longitudinal experiment to test our hypotheses (see figure 19.1). There were three experimental conditions: (1) some participants received empathy-building messages, (2) participants in the *active* control group received messages related to self-esteem or remaining objective, and (3) participants in the *passive* control group did not receive any intervention messages. Because of our small sample size, the randomization script oversampled to the empathy condition ($N = 37$) and undersampled to the other two conditions so that we could collapse them into a single control group (active control group: $N = 21$; passive control group: $N = 25$). In an ideal larger study, we would disentangle the self-esteem from the objective messages in the active control group, and we would also have enough power to examine whether there are significant differences between the two control groups (i.e. active versus passive). However, for the purpose of this initial study, we tested the difference between the empathy messages versus the *combined* control conditions.

The text messages were sent in a randomized order during blocks of time that matched participants' preferences. The messages were sent every two hours, so that a sample daily message schedule for a late-rising undergraduate might be 11 a.m., 1 p.m., 3 p.m., 5 p.m., 7 p.m., and 9 p.m. All participants, regardless of condition, also received six daily text messages that queried them on their current mood and feelings of social connection. These *query messages* were sent five minutes after the intervention text messages to those in the empathy and active control group; and on

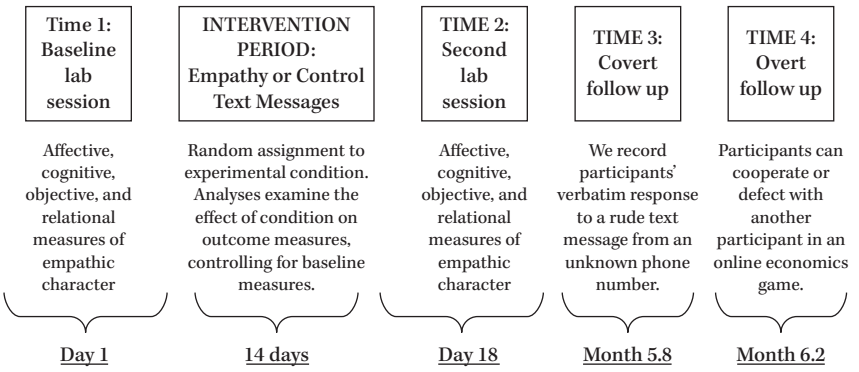
Time 1: Baseline lab session
INTERVENTION PERIOD: Empathy or Control Text Messages
TIME 2: Second lab session
TIME 3: Covert follow up
TIME 4: Overt follow up
Affective, cognitive, objective, and relational measures of empathic character
Random assignment to experimental condition. Analyses examine the effect of condition on outcome measures, controlling for baseline measures.
Affective, cognitive, objective, and relational measures of empathic character
We record participants' verbatim response to a rude text message from an unknown phone number.
Participants can cooperate or defect with another participant in an online economics game.
Day 1
14 days
Day 18
Month 5.8
Month 6.2

FIGURE 19.1 Overview of our study.

the hour to the passive control group participants. These data have not yet been analyzed.

Participants responded to an advertisement for a social-skills training study. The baseline sample included 90 college students, who were relatively gender balanced and ethnically diverse (mean age = 21; 60% female; 52% Caucasian; 33% Asian; 11% African American; and 4% other). Research assistants conducted pre- and post-laboratory assessments and were blind to participants' experimental condition.

The study involved multiple assessments at different times, as briefly described in what follows:

Time 1 and 2 measures of empathic character. During the baseline and post-intervention sessions, participants completed a number of measures of empathic character, encompassing four different categories: affective, cognitive, relational, and objective measures.

Affective. We assessed *feelings of compassion* (e.g., sympathetic, compassionate) for a stranger in distress (Batson, 2011; Batson, et al., 1997; Davis, 1983a), presented via a radio clip of one distressed person at Time 1 and via a video clip of a different distressed person at Time 2. To assess *prosocial motives for volunteer behavior*, we asked participants to "list the main reasons why you participate in unpaid volunteer work (or why you would do this, if you are not currently a volunteer)." We coded for the presence and absence of more altruistic motives (e.g., volunteering to help others) and less altruistic motives (e.g., volunteering to feel good). *Empathic concern* was assessed using the Interpersonal Reactivity Index subscale (Davis, 1983c).

Cognitive. *Perspective taking* was assessed using the Interpersonal Reactivity Index subscale (Davis, 1983c). Empathy-relevant moral beliefs were assessed via the moral *principle of care* scale (e.g., "Everybody in this world has a responsibility to help others when they need assistance"; Wilhelm & Bekkers, 2010). Beliefs in the *acceptability of aggression* were measured via the Normative Beliefs about Aggression Scale (Huesmann & Guerra, 1997). A sample item is "Suppose a young man says something bad to another young man, John. Do you think it's OK for John to scream at him?" Higher numbers indicated *lower* endorsement of aggressive responses.

Relational. *Support giving* was assessed by asking if participants had provided target groups (friends, siblings, parents, and others) with emotional support in the past week. *Support receiving* was assessed with similar questions, referring to receiving rather than giving support. The main outcome measure was the ratio of giving to receiving support.

Objective measures. We assessed *observer-rated empathy* based on research assistant ratings. After interacting with participants, research assistants were asked to rate how empathetic they thought the participant was. We also asked participants to give written responses to three *empathy-relevant scenarios*. For example, in one scenario a friend studied very hard for a test but still received

a C-minus. Participants were asked to report what they would do or say in such a situation. We coded the responses for empathy-related behaviors or emotional responses.

We assessed *baseline (Time 1) prosocial behavior* based on a standardized helping task (Batson, 2011; Batson, et al., 1997; Davis, 1983a). At the end of the laboratory session, all participants listened to a short radio program about a young college student, Katie Banks, whose parents had recently died in a car accident, and who now risks dropping out of college because of limited resources to take care of her siblings. Participants were asked to imagine her perspective while listening to the program. They were then given an opportunity to volunteer to help her with fundraising tasks by sending a sealed envelope directly to the local radio station that was helping her. Research assistants were blind to participants' helping behavior.

We assessed *Time 2 prosocial behavior* based on a similar paradigm. However, this time participants ($N = 83$) were asked to watch a short video depicting Karen Klein, a school bus monitor who was bullied by adolescent males on her school bus. This was a real video that was recorded by the actual bullies and then later posted on YouTube. After viewing the video while taking her perspective, participants were given information on an anti-bullying organization, and were then given an opportunity to volunteer for the organization. Again, responses were placed in sealed envelopes to be sent directly to the organization.

Time 3 prosocial behavior. At Time 3, which occurred on average 5.8 months after the baseline session, participants were sent a provoking text message from an unknown person. In reality, this was from a member of our research team whose phone number did not come up in Internet searches. The message was: "*stop txting me u jerk!*" Forty-three participants responded to it, half of them within the first five minutes. We recorded the responses verbatim, and then two members of the research team coded them as aggressive, neutral, or prosocial. Intercoder reliability was high ($\alpha = .95$).

Time 4 prosocial behavior. At Time 4 ($N = 60$), which occurred on average 6.2 months after the baseline session, participants played an online social dilemma game with another participant in the study. They were told that they would have a chance to earn additional money as a bonus (between $0 and $10) for playing an economics game (see table 19.2). After giving them general

TABLE 19.2 } **Decision chart for social dilemma game**

	YOU: Cooperate	**YOU: Defect**
YOUR PARTNER: Cooperates	You get $6 bonus Your partner gets $6 bonus	You get $10 bonus Your partner $0 gets bonus
YOUR PARTNER: Defects	You get $0 bonus Your partner gets $10 bonus	You get $0 bonus Your partner gets $0 bonus

background information about economics games, they were then given the specific instruction to look at the following table and make a decision to either cooperate or defect. Prior research has found that people who are randomly assigned to take the perspective of their partner are more likely to cooperate with that specific person (Batson & Moran, 1999), but at the expense of contributing to the collective good (Batson et al., 1995).

NOTES ON STUDY IMPLEMENTATION

Before presenting the results, it is worth briefly noting some of the obstacles we encountered while trying to collect these data. These are currently barriers to the wider use of such methods, but I hope that this will simply be a historical note at some point in the future. Although we only collected data from 90 participants, rendering it more of a pilot study, it was extremely resource intensive. Besides the principal investigator, it required one full-time staff member, one computer programmer, and 11 undergraduate research assistants. Scheduling the pre- and post-sessions separated by a two-week period proved to be difficult with busy college students, and our follow-up success rate seemed to be determined by the time of term, with most dropouts occurring at the end of the academic term. Yet this difficulty is to be expected with pre- and post-experimental designs.

The major problem that we encountered during this study was the lack of adequate available tools to help administer research studies using mobile technology. This is a problem if researchers want to more widely use this method. After doing some research on potential programs, we ended up using RedOxygen, an Internet-based program that is intended to be used to send a single message to a large group of cell phone numbers (i.e., spam). We needed to do something different from what the program was designed for: to send and receive a large group of messages to a single cell phone each day for a 14-day period. To accomplish our goal, we had to create a computer program that first randomly assigned participants to one of the conditions, then created a randomized order of the text messages for each participant, and then created a customized delivery schedule for each participant. When we ran the program, it logged into RedOxygen and sent the RedOxygen server separate requests for each text message that needed to be sent. This involved sending 168 text messages per participant for the active conditions (six intervention messages and six query messages for 14 days). A separate issue was that the program was not designed for data collection (most spam messengers do not expect responses). Although the program was able to collect the responses, these were spit out as a large text file and took intensive processing to be made usable data for analyses.

At the time we were running the study, there was no known tool that was easy for researchers to use (at the level of Qualtrics, for example) and also had

the capability of randomizing participants into experimental conditions. There are a number of studies in public health that use text messaging; however, most of these required at least some programming. I now know of a few available programs that are suitable for ecological momentary assessment purposes (i.e., data collection in real time), but only one that has begun beta testing a more user-friendly program for ecological momentary interventions (SurveySignal). These programs are described and compared in my recent chapter on mobile-based psychosocial interventions (see Konrath, 2015).

Our research team needed to keep track of a large volume of text messages that were both directed to and received from participants, estimated at approximately 20,000 messages in total (not accounting for dropout or missed responses) over the initial six-month data collection period. Each message cost 10 cents, and messages are purchased in large batches. When mobile users fail to receive a spam message, they are unlikely to notice or complain. However, the combination of highly motivated participants and less-than-reliable implementation software meant that a lot of time was spent carefully tracking and double-checking the messages: whether they were actually sent, whether participants actually received them, and whether participants were responding to them. We were careful to respond in a timely fashion whenever there were problems with message delivery, to ensure that all participants received the same dosage.

One other point of consideration when running these types of studies is how to properly compensate participants for a study that involved two lab visits and 14 days of intensive involvement. We decided on a compensation strategy in which participants received a larger remuneration on the second lab visit than on the first, to thank them for staying in the study. They also received a small payment each time they responded to a query message and a daily bonus if they responded to all six query messages in a single day. We deliberately did not incentivize the six daily intervention messages in order to limit participants' extrinsic motivation to follow the directions.

Study Results and Interpretation

ENGAGEMENT METRICS AND DATA ANALYSES

The study's Time 2 dropout rate was less than 8%, with 83 of the original 90 participants returning to the lab for the second time. Of those 83 participants, 95% (N = 79) agreed to be contacted by the research team in the future, and of those 79, 43 participants (54%) responded to the hostile text message from the unknown number (Time 3; covert follow-up). Of the 79 participants who agreed to be contacted in the future, 60 participants (76%) completed the online follow-up survey in which they played a social dilemma game (Time 4; overt follow-up). On average, included participants responded to 78.2 of

the 84 (93%) incentivized query messages that they received across the 14-day period. There was no effect of condition on the number of responses to the query messages. Taken together, these metrics indicate that participants were highly engaged with the program.

To analyze the data, we examined the effect of condition (empathy versus the collapsed control conditions) and gender on each of the outcomes, controlling for their baseline scores whenever that was possible. The use of random assignment and the statistical control for baseline scores gives us more confidence that the outcomes we found were not because of differences between the groups at baseline.

RESULTS

As can be seen in table 19.3, the question of whether our intervention affected empathic character has a complex answer. Participants who received the empathy-building text messages reported more other-oriented motives for volunteering (i.e., to help) and less self-oriented ones (i.e., to feel good), less beliefs that aggression is an acceptable response (males), and more giving than receiving of social support (males). They also offered more hours of help and had lower feelings of personal distress when exposed to a needy person. Personal distress feelings are typically seen as self-oriented emotional responses, rather than empathic ones (Batson, 2011).

In addition, participants in the empathy condition were rated by observers as higher in empathy, showed more emotional empathy in imagined scenarios (e.g., emotional resonance, acknowledging emotion), and responded more prosocially to a stranger who sent them an unprovoked hostile text message six months after the intervention. To us, this response is notable, because the hostile text message was not obviously linked to the study, and it was sent so long after the intervention that it is doubtful that participants were even thinking of the study when they responded.

However, not all of the results were so straightforward. Those in the empathy-building condition also saw themselves as lower in two traits that I consider to be important indicators of empathic character: dispositional empathic concern and the moral principle of care (males). In addition, although they showed more *emotional* empathy in the imagined scenarios, they simultaneously showed less *practical* empathy. So, for example, when they imagined a friend coming to them after a breakup, they were more likely to say that they would acknowledge the friend's emotion, but less likely to say that they would offer practical help. Moreover, the empathy-building condition had no effect on dispositional perspective taking, feelings of compassion for a needy person, and the tendency to cooperate with other participants in a social dilemma game on the overt follow-up survey six months later.

TABLE 19.3 } **Summary of results**

Time 2	Result	Direction of effect	Type
SELF-REPORTED MEASURES			
Motives for volunteering behavior: To help others	↑ in empathy condition	+	Motive
Motives for volunteering: To feel good	↓ in empathy condition	+	Motive
Beliefs that aggression is acceptable	↓ in empathy condition among males	+	Belief
Ratio of giving to receiving social support	↑ in empathy condition among males	+	Self-report behavior
IRI Perspective taking	⊗ no effect of condition	0	Trait
IRI Empathic concern	↓ in empathy condition	–	Trait
Moral principle of care	↓ in empathy condition among males	–	Trait
PROSOCIAL BEHAVIORS AND EMOTIONS			
Observer rated empathy	↑ in empathy condition (marginal)	+	Observable Behavior
Feelings of personal distress in response to needy target	↓ in empathy condition	+	Emotional response
Feelings of compassion in response to needy target	⊗ no effect of condition	0	Emotional response
Helping behavior in response to needy target	↑ ***hours*** of help in empathy condition (marginal); no effect on ***decision to help***	+	Observable Behavior
Emotional empathy in imagined scenarios (e.g. emotional resonance, acknowledging emotion)	↑ in empathy condition	+	Emotional response
Practical empathy in imagined scenarios (e.g. offering to listen or help)	↓ in empathy condition (marginal)	–	Self-report behavior
TIME 3			
Response to hostile text message (higher = more prosocial response)	↑ in empathy condition	+	Observable Behavior
TIME 4			
Cooperation in social dilemma game	⊗ no effect of condition	0	Observable Behavior

Note. Baseline scores were added as covariates whenever possible. For direction of effect, a plus sign (+) indicates expected, a minus sign (–) indicates opposite of what was expected, and a zero (0) indicates a null result.

Overall, on the 15 key measures of empathic character, nine results were in the expected direction, three results were null (possibly because of low power), and three results were in the *opposite* direction. A few of the results in the expected direction were only statistically marginal, which may again be because of low statistical power. In addition, some outcomes depended on gender and others did not, and there did not seem to be a clear pattern that could explain these results except that the intervention seemed to be more effective for males. Overall, the evidence that the empathy-building text messages increased participants' empathic character is not perfectly clear, but is intriguing for a study with this relatively small sample size.

REFLECTIONS ON THE EFFICACY OF THE EMPATHY-BUILDING INTERVENTION

We are left with a question that is similar to those the chapter began with: can we really teach old dogs new tricks, can leopards change their spots, and do old habits die hard? Did we really change participants' ingrained habits of empathic character after only two weeks? The answer is messier than may be desired, and I think this has to do with our definition and measurement of empathic character. We also need to keep in mind that because of the relatively small sample size, the power to detect significant differences is limited.

Still, some overarching patterns can be identified if we carefully examine the results. In table 19.4, I attempt to do so by grouping the measures into different categories: observable behaviors, motives, emotional responses, beliefs, self-reported behaviors, and traits. I gave each effect that was in the expected direction a score of 1, each null result a score of 0, and each effect that was in the opposite direction a score of –1. When the data is examined from this perspective, it appears that the results depend on the type of measure that we used to assess empathic character.

The empathy-building text messages changed observable behaviors most successfully. Participants who received the empathy messages were rated by others as more empathic, offered more time to a person in distress, and

TABLE 19.4 } **Change in empathic character depends upon how it is measured**

Type of measure	Direction of effect	Total score
Observable behaviors	+ + + 0	3
Motives	+ +	2
Emotional responses	+ + 0	2
Beliefs	+	1
Self-report behaviors	+ –	0
Traits	– – 0	–2

Note. Plus sign (+) is coded as 1, null sign (0) is coded as 0, and minus sign (–) is coded as –1.

responded more prosocially after being provoked by a hostile text message six months later. There was no effect on their cooperation with another unknown participant in a social dilemma game, yet prior research finds that when people are asked to empathize with a partner, they are more likely to cooperate (Batson & Moran, 1999). Our study does not replicate this finding, and it is unclear why this is the case. It is possible that methodological differences mattered. In prior research, the partner was a person of the same gender and academic major who was supposedly playing the game with participants in a different room. In our study, the partner was less concrete. We told participants that we would randomly match them with another participant who completed the survey either right before or right after them. It is possible that in our study the partner was too psychologically distant to have an emotional impact.

Typically, behavioral outcomes are seen as strong evidence in favor of an intervention compared to other less observable outcomes. But in the case of character traits, behaviors in themselves do not give us enough information. People can act in kind ways for many reasons, including that it makes them look good or makes it more likely that they will get what they want. However, the empathy text messages also changed participants' motives for volunteering (see tables 19.3 and 19.4). Most of the students in the study regularly volunteered for nonprofit organizations, and the two-week intervention affected their stated reasons for volunteering, even when controlling for their baseline motives.

The empathy condition also affected emotional aspects of empathy. Participants felt less personal distress when exposed to a needy target, and also imagined that they would experience more emotional empathy when presented with hypothetical scenarios. Yet the empathy condition did not increase participants' feelings of compassion for the needy target. This might have been because all participants were asked to *imagine the target's perspective* while listening to a radio program (Time 1) or watching a video (Time 2). In hindsight, it would have been a better idea to let participants naturally vary in how they responded to the stimuli, but at the time, our rationale was to keep participants' mindset constant while consuming the media clips. Unfortunately, in doing so, we may have increased feelings of compassion to such an extent that any differences between the two experimental groups were erased.

The results for beliefs and self-report behaviors were inconsistent. In terms of beliefs, male participants in the empathy condition were less likely to believe that aggression was an acceptable response, and in terms of self-report behaviors, they reported more giving than receiving transactions in their close social relationships. However, at the same time, those in the empathy condition imagined that they would be *less* likely to offer practical assistance when posed with hypothetical scenarios about someone in need. Instead, they were more likely to offer subtler and more complex emotional responses in these scenarios. This might make sense if we posit that empathic character moves people beyond an obvious checklist of helping behaviors, and more into the

imaginative space of people who are in need. This explanation also fits with some of the other results finding that the empathy messages led to more subtle empathic acts. For example, participants in both conditions were *equally likely to help* the needy target, but those in the empathy condition offered *more hours of help* than those in the control conditions. So, rather than simply helping or saying that one would help in an imagined scenario, empathic character may increase the degree of helping (e.g., the number of hours) and the emotional subtlety of it (e.g., feeling with others, acknowledging their emotions).

The most interesting and unexpected finding was that participants who received empathy-building messages scored *lower* in trait empathic concern and the moral principle of care (males). This went against our predictions, and we can only speculate on potential explanations. First, it is possible that these lower self-perceptions stem from contrast effects (Schwarz & Bless, 1992a, 1992b). For example, it is possible that the frequent daily exposure to the high standard of empathy and compassion made participants see themselves as low in empathy and the moral principle to care, even if their behaviors indicated otherwise. At times they may have found the instructions annoying or difficult to perform, and if so, this would have made them conclude that they are not very empathic people. Some research suggests that it is possible to have such lowered self-perceptions even in the presence of increased trait-relevant behaviors. One study, for example, found that when participants were asked to recall 12 instances of their assertiveness, they saw themselves as *less* assertive than those who were asked to recall six instances, despite the fact that the first group of participants actually recalled *more* assertive behaviors (Schwarz et al., 1991). Maybe the intensive daily reminders to focus on others made participants in the empathy condition more aware of how they deviate from this high standard.

Second, these unexpected lower self-perceptions in the empathy condition could be explained by frame-of-reference effects (Heine, Lehman, Peng, & Greenholtz, 2002; Schmit, Ryan, Stierwalt, & Powell, 1995; Wood, Brown, Maltby, & Watkinson, 2012). Perhaps the frequent daily empathy messages made participants shift their beliefs about how empathic most other people are. Participants may have assumed that most people are pretty empathic, habitually focusing on others most of the time each day. If so, participants could plausibly see themselves as less empathic in comparison.

It is worth noting that although these lowered self-perceptions were unexpected, they help to rule out social desirability as an alternative explanation for many of our results. After receiving six messages per day telling them to think about and feel for others, it is likely that participants guessed the true purpose of our study. Given this, some might argue that our results can best be explained by their desire to look good to researchers or to themselves. Yet we can rule out such an explanation both because of the low self-perceptions of empathic concern, and also because participants responded more prosocially to a hostile provocation that did not appear to be connected to the study.

Future Directions

Although we did not have entirely consistent results, taken together, the evidence suggests that we were able to teach old dogs at least *some* new tricks, change a few "spots" on our participants, and at least begin to form new empathic habits among them. This work is only at the beginning stages though, with many potential future directions to pursue.

First, in this same dataset, we also assessed psychological well-being and mental health outcomes, and we will be analyzing these data in the months to come. This can help to more clearly determine the causal role of empathy in health and well-being, since most of the research to date is based on correlational or longitudinal data (Konrath & Brown, 2012). Although Batson and his colleagues have convincingly demonstrated that being randomly assigned to an empathic mindset increases altruistically motivated prosocial behaviors (Batson, 2011), we know little about the causal relationship between empathy and psychological well-being, mental health, physical health, relationship outcomes, and violence and aggression (Konrath & Grynberg, 2013). This empathy-building text message program can thus help to answer a number of important research questions.

Next, we are currently designing future studies with larger and more diverse participant samples. Increasing the sample size will allow us to disentangle the effectiveness of the program relative to different kinds of control groups (e.g., self-esteem vs. objective vs. passive control groups). Expanding this research outside of the domain of privileged college students is another necessary next step. We are currently working with our technology partners (*HopeLab* and *Textizen*) to test this program on a large and diverse sample of adolescents. There are a number of other participant groups who could also likely benefit from this program or an adapted version of it: new parents, who need help prioritizing the needs of their infants; medical students, who have been shown to decline in empathy during medical school (Hojat et al., 2009; Neumann et al., 2011); teachers, whose empathy scores are linked to their students' objective performance (Chang, et al., 1981; Waxman, 1983); and other human service professionals.

Given that participants in the empathy condition saw themselves as less empathic compared to controls, we are also testing dosage levels in our future research. Since it is possible that receiving so many messages per day backfired in terms of self-perceptions, we will be running studies to determine the optimal number of messages per day. For example, we are asking whether we will get similar effects if participants receive fewer messages per day. Indeed, we are even testing whether one message per day would be sufficient for character change to occur.

Finally, we are also exploring other mobile-based methods for building empathic character, with support from the *John Templeton Foundation*. The current text messaging program uses a didactic approach to building habits, but such overt instructions could have the unintended side effect of increasing

reactance or disinterest among some people. Thus, we are also creating a game-based empathy-building smartphone app. Our hope is that the game play is so interesting and enjoyable that developing a set of empathy-related skills is seen as a bonus outcome, rather than the main goal, of playing. This philosophy is key to the emerging serious games movement in public health and psychology (Baldwin & Dandeneau, 2009).

In future work, I also hope to examine whether the duration of exposure to the empathy-building program matters. In the current study, participants received the empathy-building messages for a two-week period. However, in a recent review of the literature, I found that most physical health text-messaging interventions were six weeks or longer, with some lasting up to 14 months (Konrath, 2015). Although we found that participants responded more prosocially to a hostile text message approximately six months after receiving the empathy program, two weeks may not be long enough to change a habit on a long term basis. How long should a character change program be? One recent study examined how long it took for a new health behavior to feel automatic, which is one way of characterizing a habit (Lally, Van Jaarsveld, Potts, & Wardle, 2010). Participants were asked to carry out a new health behavior of their choice for a period of 84 days (e.g., drink a glass of water with lunch). On average, it took 66 days for automaticity scores to plateau. Considering this, interventions that are nine weeks or more might be more effective at building habit-like character traits than shorter ones.

Conclusion

This chapter asked whether text messages could make people kinder, but I also hope that readers will see the larger potential of mobile-based interventions in the study of character. Can text messages also make people more honest, more optimistic and hopeful, and more grateful? The current study has implications for these and other character traits, and I recommend that other scholars consider whether it would be useful to develop and test tools that could both build valued character traits, and be used as scientific tools to test hypotheses about those traits. It is currently very difficult to run such studies, and there is a need for user-friendly tools to more easily implement mobile-based interventions in the future.[1]

[1] This chapter was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation via grant number 47993. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. This project was also supported by the Office of the Vice President for Research at the University of Michigan. Thank you to Andrea Fuhrel-Forbis for coordinating the day-to-day aspects of this project, Matthew O'Donnell for assistance with programming, and undergraduate members of the Interdisciplinary Program for Empathy and

References

Abelson, J. L., Erickson, T. M., Mayer, S., Crocker, J., Briggs, H., Lopez-Duran, N. L., & Liberzon, I. (in press). Brief cognitive intervention can modulate neuroendocrine stress responses to the Trier Social Stress Test: Buffering effects of a compassionate goal orientation. *Psychoneuroendocrinology*, 44, 60–70.

Abraham, A., Pocheptsova, A., & Ferraro, R. (2012). *The effect of mobile phone use on prosocial behavior*. Unpublished manuscript.

Adams, A. (2010). *The relationship among illness representations, risk representations, empathy, and preventive health behaviors.* Psy.diss., Marywood University.

Amato, P. R. (1990). Personality and social network involvement as predictors of helping behavior in everyday life. *Social Psychology Quarterly*, *53*(1), 31–43.

Aspy, D. N., & Roebuck, F. N. (1972). An investigation of the relationship between student levels of cognitive functioning and the teacher's classroom behavior. *Journal of Educational Research*, *65*(8), 365–368.

Au, A. M. L., Wong, A. S. K., Lai, M. K., & Chan, C. C. H. (2011). Empathy, coping, social support, and mental health in local and migrant adolescents in Beijing. *International Journal on Disability and Human Development*, *10*(3), 173–178. doi:10.1515/ijdhd.2011.030.

Baldwin, M. W., & Dandeneau, S. D. (2009). Putting social psychology into serious games. *Social and Personality Psychology Compass*, *3*(4), 547–565.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54*(1), 39–57.

Bar-Tal, D., & Raviv, A. (1979). Consistency of helping-behavior measures. *Child Development*, *50*(4), 1235–1238.

Barak, A., Engle, C., Katzir, L., & Fisher, W. A. (1987). Increasing the level of empathic understanding by means of a game. *Simulation & Gaming*, *18*(4), 458–470.

Barkai, J., & Fine, V. (1983). Empathy training for lawyers and students. *Southwestern University Law Review*, *13*, 505–529.

Barnett, M. A. (1984). Similarity of experience and empathy in preschoolers. *The Journal of Genetic Psychology*, *145*(2), 241–250.

Barnett, M. A., Howard, J. A., Melton, E. M., & Dino, G. A. (1982). Effect of inducing sadness about self or other on helping behavior in high-and low-empathic children. *Child Development*, *53*(4), 920–923.

Baron, R. A. (1997). The sweet smell of . . . helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls. *Personality and Social Psychology Bulletin*, *23*, 498–503.

Batson, C. D. (2011). *Altruism in humans.* New York, NY Oxford University Press.

Batson, C. D., Batson, J. G., Todd, R. M., Brummett, B. H., Shaw, L. L., & Aldeguer, C. M. (1995). Empathy and the collective good: Caring for one of the others in a social dilemma. *Journal of Personality and Social Psychology*, *68*(4), 619.

Altruism Research for assistance with data collection: Erin Ackert, Mahnoor Asghar, Robyn Bernstein, Carren Cheng, Curtis Hill, Melissa Juganda, Muneer Khalid, Boyoung Kim, Grace Kirkley, Jon Von Leon Lee, and Rita Morris. I thank Christian Miller for the character traits that drove him to expertly lead The Character Project grantees through three years of rich intellectual development. I gratefully acknowledge the comments of William Fleeson, Christian Miller, and other members of the Character Project for their comments on this chapter and the conference presentations on which it was based.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and Social Psychology Bulletin*, *23*(7), 751–758.

Batson, C. D., & Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *European Journal of Social Psychology*, *29*(7), 909–924.

Beck, R. S., Daughtridge, R., & Sloane, P. D. (2002). Physician-patient communication in the primary care office: a systematic review. *The Journal of the American Board of Family Practice*, *15*(1), 25–38.

Bekkers, R. (2005). Participation in voluntary associations: Relations with resources, personality, and political values. *Political Psychology*, *26*(3), 439–454.

Bekkers, R. (2006). Traditional and health-related philanthropy: The role of resources and personality. *Social psychology quarterly*, *69*(4), 349–366.

Bierhoff, H. W., Klein, R., & Kramp, P. (1991). Evidence for the altruistic personality from data on accident research. *Journal of Personality*, *59*(2), 263–280.

Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, *9*(1–2), 52–69.

Brehm, S. S., Fletcher, B. L., & West, V. (1981). Effects of empathy instructions on first-grader's liking of other people. *Child Study Journal*, *11*(1), 1–15.

Bryan, J. H., & Walbek, N. H. (1970). Preaching and practicing generosity: Children's actions and reactions. *Child Development*, *41*(2), 329–353.

Burtson, P. L., & Stichler, J. F. (2010). Nursing work environment and nurse caring: relationship among motivational factors. *Journal of Advanced Nursing*, *66*(8), 1819–1831. doi: 10.1111/j.1365-2648.2010.05336.x.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology*, *50*(3), 559.

Carlo, G., Okun, M. A., Knight, G. P., & de Guzman, M. R. T. (2005). The interplay of traits and motives on volunteering: agreeableness, extraversion and prosocial value motivation. *Personality and Individual Differences*, *38*(6), 1293–1305.

Carpenter, C. J. (2012). Narcissism on Facebook: Self-promotional and anti-social behavior. *Personality and Individual Differences*, *52*(4), 482–486. doi.org/10.1016/j.paid.2011.11.011.

Chang, A. F., Berger, S. E., & Chang, B. (1981). The relationship of student self-esteem and teacher empathy to classroom learning. *Psychology: A Journal of Human Behavior*, *18*(4), 21–25.

Clary, E. G., Snyder, M., & Stukas, A. A. (1996). Volunteers' motivations: Findings from a national survey. *Nonprofit and Voluntary Sector Quarterly*, *25*(4), 485–505.

Coffman, S. L. (1981). Empathy as a relevant instructor variable in the experiential classroom. *Group & Organization Management*, *6*(1), 114–120.

Cole-Lewis, H., & Kershaw, T. (2010). Text messaging as a tool for behavior change in disease prevention and management. *Epidemiologic Reviews*, *32*(1), 56–69.

Coyne, S. M., Stockdale, L., Busby, D., Iverson, B., & Grant, D. M. (2011). "I luv u:)!": A descriptive study of the media use of individuals in romantic relationships. *Family Relations*, *60*(2), 150–162.

Darley, J. M., & Batson, C. D. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*(1), 100–108.

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85. Available online at: http://skynet.eckerd.edu/academics/psychology/files/Davis_1980.pdf

Davis, M. (1983a). The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. *Journal of Personality*, *51*(2), 167–184.

Davis, M. (1983b). Empathic concern and the Muscular Dystrophy Telethon: Empathy as a multidimensional construct. *Personality and Social Psychology Bulletin*, *9*(2), 223–229.

Davis, M. (1983c). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. doi:10.1037/0022-3514.44.1.113

Davis, M., & Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *Journal of Research in Personality*, *25*(1), 70–87.

Davis, M., & Kraus, L. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy, 144–168*. New York, NY: Guilford.

Davis, M., Luce, C., & Kraus, S. J. (1994). The heritability of characteristics associated with dispositional empathy. *Journal of Personality*, *62*(3), 369–391.

Derksen, F., Bensing, J., & Lagro-Janssen, A. (2013). Effectiveness of empathy in general practice: a systematic review. *British Journal of General Practice*, *63*(606), e76-e84.

Dey, A. K., Wac, K., Ferreira, D., Tassini, K., Hong, J.-H., & Rojas, J. (2011). Getting closer: an empirical investigation of the proximity of user to their smart phones. *Ubicomp '11*, *Proceedings of the 13th International Conference on Ubiquitous computing*, 163–172.

Dillon, M., & Wink, P. (2007). *In the course of a lifetime: Tracing religious belief, practice, and change* Berkeley and Los Angeles: University of California Press.

Dyrbye, L. N., Massie, F. S., Eacker, A., Harper, W., Power, D., Durning, S. J., . . . Shanafelt, T. D. (2010). Relationship between burnout and professional conduct and attitudes among US medical students. *JAMA: The Journal of the American Medical Association*, *304*(11), 1173–1180. doi:10.1001/jama.2010.1318.

Eisenberg, N., Carlo, G., Murphy, B., & Court, P. (1995). Prosocial development in late adolescence: A longitudinal study. *Child Development*, *66*(4), 1179–1197.

Eisenberg, N., Guthrie, I. K., Cumberland, A., Murphy, B. C., Shepard, S. A., Zhou, Q., & Carlo, G. (2002). Prosocial development in early adulthood: a longitudinal study. *Journal of Personality and Social Psychology*, *82*(6), 993.

Eisenberg, N., Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development*, *70*(6), 1360–1372.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*(1), 91.

Eisenberg, N., Miller, P. A., Schaller, M., Fabes, R. A., Fultz, J., Shell, R., & Shea, C. L. (1989). The role of sympathy and altruistic personality traits in helping: A reexamination. *Journal of Personality*, *57*(1), 41–67.

Eisenberg, N., Shell, R., Pasternack, J., Lennon, R., Beller, R., & Mathy, R. M. (1987). Prosocial development in middle childhood: A longitudinal study. *Developmental Psychology*, *23*(5), 712.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In Ericsson, K.A., Charness, N., Feltovich, P.,

& Hoffman, P. (Eds.) *The Cambridge Handbook of Expertise and Expert Performance*, 683–703, New York: Cambridge University Press.

Fagin-Jones, S., & Midlarsky, E. (2007). Courageous altruism: Personal and situational correlates of rescue during the Holocaust. *Journal of Positive Psychology*, *2*(2), 136–147.

Faulkner, X., & Culwin, F. (2005). When fingers do the talking: A study of text messaging. *Interacting with Computers*, *17*(2), 167–185.

Feshbach, N. (1984). Empathy, empathy training and the regulation of aggression in elementary school children In Kaplan, R.M., Konecni, V.J., & Novaco, R.W. (Eds.) *Aggression in Children and Youth,* 192–208. The Hague/Boston/Lancaster, Martinus Nijhoff Publishers.

Feshbach, N., & Konrad, R. (2001). Modifying aggression and social prejudice *Prevention and Control of Aggression and the Impact on its Victims*, 355–360. In Martinez, M. (Ed.), Prevention and control of aggression and the impact on its victims, New York, NY: Kluwer Academic/Plenum Publishers.

Fiordelli, M., Diviani, N., & Schulz, P. J. (2013). Mapping mHealth research: a decade of evolution. *Journal of Medical Internet Research*, *15*(5), e95, DOI: 10.2196/jmir.2430

Fjeldsoe, B. S., Marshall, A. L., & Miller, Y. D. (2009). Behavior change interventions delivered by mobile telephone short-message service. *American Journal of Preventive Medicine*, *36*(2), 165.

Galinsky, A. D., Maddux, W. W., Gilin, D., & White, J. B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science*, *19*(4), 378–384.

Goldstein, T. R., & Winner, E. (2012). Enhancing empathy and theory of mind. *Journal of Cognition and Development*, *13*(1), 19–37.

Grühn, D., Rebucal, K., Diehl, M., Lumley, M., & Labouvie-Vief, G. (2008). Empathy across the adult lifespan: Longitudinal and experience-sampling findings. *Emotion*, *8*(6), 753–765.

Hartshorne, H., May, M. A., & Maller, J. (1928–1930). *Studies in the nature of character, Volumes 1-3*. New York: Macmillan Company.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903.

Herbert, L., Owen, V., Pascarella, L., & Streisand, R. (2013). Text message interventions for children and adolescents with type 1 diabetes: A systematic review. *Diabetes Technology & Therapeutics*, *15*(5), 362–370.

Ho, S. S., Konrath, S., Brown, S., & Swain, J. E. (2014). Empathy and stress related neural responses in maternal decision making. *Frontiers of Neuroscience*, *8*, 152. doi:10.3389/fnins.2014.00152.

Hodges, S. D., & Wegner, D. M. (1997). Automatic and controlled empathy. *Empathic Aaccuracy*, 311–339.

Hojat, M., Vergare, M. J., Maxwell, K., Brainard, G., Herrine, S. K., Isenberg, G. A., . . . Gonnella, J. S. (2009). The devil is in the third year: A longitudinal study of erosion of empathy in medical school. *Academic Medicine*, *84*(9), 1182–1191.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, *72*(2), 408.

Huston, T. L., Ruggiero, M., Conner, R., & Geis, G. (1981). Bystander intervention into crime: A study based on naturally-occurring episodes. *Social Psychology Quarterly*, *44*(1), 14–23.

Ickes, W. (1997). *Empathic Accuracy*. New York: Guilford Press.

Ickes, W., & Simpson, J. A. (2003). Motivational Aspects of Empathic Accuracy. In Fletcher, G. & Clark, M. *Blackwell Handbook of Social Psychology: Interpersonal Processes*, *2*, 229. Malden, MA: Blackwell Publishers Ltd.

Ironson, G. (2007). Altruism and health in HIV. In S. Post (Ed.), *Altruism and health* (pp. 70–81). New York: Oxford University Press.

Ironson, G., Solomon, G., Balbin, E., O'Cleirigh, C., George, A., Kumar, M., . . . Woods, T. (2002). The Ironson-Woods spirituality/religiousness index is associated with long survival, health behaviors, less distress, and low cortisol in people with HIV/AIDS. *Annals of Behavioral Medicine*, *24*(1), 34–48. doi:10.1207/s15324796abm2401_05.

Isen, A., Horn, N., & Rosenhan, D. (1973). Effects of success and failure on children's generosity. *Journal of Personality and Social Psychology*, *27*(2), 239–247.

Isen, A., & Levin, P. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*, *21*(3), 384–388.

Kalliopuska, M. (1992). Attitudes towards health, health behaviour, and personality factors among school students very high on empathy. *Psychological Reports*, *70*(3), 1119–1122.

Kerber, K. W. (1984). The perception of nonemergency helping situations: Costs, rewards, and the altruistic personality. *Journal of Personality*, *52*(2), 177–187.

Klein, K. J., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin*, *27*(6), 720–730.

Koestner, R., Franz, C., & Weinberger, J. (1990). The family origins of empathic concern: A 26-year longitudinal study. *Journal of Personality and Social Psychology*, *58*(4), 709.

Kohn, A. (1991). *Caring kids: The role of the schools. Phi Delta Kappan*, *72*(7), 496–506.

Konrath, S. (2012). The empathy paradox: Increasing disconnection in the age of increasing connection. In R. Luppicini (Ed.), *Handbook of research on technoself: Identity in a technological society* (pp. 204–228). Hershey, PA: IGI Global.

Konrath, S. (2014). Social support. In A. T. Scull (Ed.), *Cultural sociology of mental illness: An A-to-Z guide*: Thousand Oaks, CA: SAGE Publications, 820–823.

Konrath, S. (2015). Positive technology: Using mobile phones for psychosocial interventions. In Z. Yan (Ed.), *Encyclopedia of mobile phone behavior*. Hershey, PA: IGI Global.

Konrath, S., & Brown, S. L. (2012). The effects of giving on givers. In N. Roberts & M. Newman (Eds.), *Handbook of health and social relationships*. Washington, DC: American Psychological Association, Chapter 2.

Konrath, S., Bushman, B., & Campbell, W. K. (2006). Attenuating the link between threatened egotism and aggression. *Psychological Science*, *17*(11), 995–1001. doi:PSCI1818 [pii]10.1111/j.1467-9280.2006.01818.x.

Konrath, S., Chopik, W., Hsing, C., & O'Brien, E. (2014). Changes in attachment style in American college students over time: A meta-analysis. *Personality and Social Psychology Review*, *18*(4), 326–348..

Konrath, S., & Fuhrel-Forbis, A. (2011a). Motives for volunteering predict volunteering regularity and frequency. *Unpublished data*.

Konrath, S., & Fuhrel-Forbis, A. (2011b). Self-rated caring traits are associated with lower mortality in older adults. *Unpublished data*.

Konrath, S., & Grynberg, D. (2013). The positive (and negative) psychology of empathy. In D. Watt & J. Panksepp (Eds.), *The neurobiology and psychology of empathy*: Nova Science Publishers In press.

Konrath, S., O'Brien, E., & Hsing, C. (2011). Changes in dispositional empathy in American college students over time: a meta-analysis. *Personality and Social Psychology Review*, *15*(2), 180–198. doi:1088868310377395 [pii]. 10.1177/1088868310377395.

Krebs, D., & Miller, D. T. (1985). Altruism and aggression. *Handbook of Social Psychology*, *2*, 1–71.

Krebs, D., & Sturrup, B. (1982). Role-taking ability and altruistic behaviour in elementary school children. *Journal of Moral Education*, *11*(2), 94–100.

Kremer, J. F., & Dietzen, L. L. (1991). Two approaches to teaching accurate empathy to undergraduates: Teacher-intensive and self-directed. *Journal of College Student Development*, *32*(1), 69–75.

Krevans, J., & Gibbs, J. C. (1996). Parents' use of inductive discipline: Relations to children's empathy and prosocial behavior. *Child Development*, *67*(6), 3263–3277.

Krishna, S., Boren, S. A., & Balas, E. A. (2009). Healthcare via cell phones: A systematic review. *Telemedicine and e-Health*, *15*(3), 231–240.

Kurtz, R. R., & Grummon, D. L. (1972). Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *Journal of Consulting and Clinical Psychology*, *39*(1), 106.

Lally, P., Van Jaarsveld, C. H., Potts, H. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009.

Larrieu, J., & Mussen, P. (1986). Some personality and motivational correlates of children's prosocial behavior. *Journal of Genetic Psychology*, *147*(4), 529–542.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, *89*(2), 308.

Levitas, D. (2013). Always connected: How Smartphones and Social Keep Us Engaged. *International Data Corporation, Sponsored by Facebook*. Retrieved from http://www.nu.nl/files/IDC-Facebook%20Always%20Connected%20%281%29.pdf on March 9, 2015.

Liang, X., Wang, Q., Yang, X., Cao, J., Chen, J., Mo, X., . . . Gu, D. (2011). Effect of mobile phone intervention for diabetes on glycaemic control: A meta-analysis. *Diabetic Medicine*, *28*(4), 455–463.

Lobchuk, M. M., & Bokhari, S. A. (2008). *Linkages among empathic behaviors, physical symptoms, and psychological distress in patients with ovarian cancer: A pilot study. Oncology Nursing Forum, 35(5),* 808–814..

Locraft, C., & Teglasi, H. (1997). Teacher rated empathic behaviors and children's TAT stories. *Journal of School Psychology*, *35*(2), 217–237.

Matthews, K. A., Batson, C. D., Horn, J., & Rosenman, R. H. (1981). "Principles in his nature which interest him in the fortune of others . . .": The heritability of empathic concern for others1. *Journal of Personality*, *49*(3), 237–247.

McClelland, D., & Krishnit, C. (1988). The effect of motivational arousal through films on salivary immunoglobulin A. *Psychology & Health*, *2*(1), 31–52.

McNeely, B. L., & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology*, *79*(6), 836.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy1. *Journal of Personality*, *40*(4), 525–543.

Midlarsky, E., Fagin Jones, S., & Corley, R. P. (2005). Personality correlates of heroic rescue during the Holocaust. *Journal of Personality*, *73*(4), 907–934.

Militello, L. K., Kelly, S. A., & Melnyk, B. M. (2012). Systematic review of text-messaging interventions to promote healthy behaviors in pediatric and Adolescent populations: Implications for clinical practice and research. *Worldviews on Evidence-Based Nursing*, *9*(2), 66–77.

Miller, C. (2013). *Moral character: An empirical theory*. Oxford: Oxford University Press.

Milton, J. G., Small, S. S., & Solodkin, A. (2004). On the road to automatic: Dynamic aspects in the development of expertise. *Journal of Clinical Neurophysiology*, *21*(3), 134–143.

Neumann, M., Edelhäuser, F., Tauschel, D., Fischer, M. R., Wirtz, M., Woopen, C., . . . Scheffer, C. (2011). Empathy decline and its reasons: A systematic review of studies with medical students and residents. *Academic Medicine*, *86*(8), 996–1009.

Nielsen. (2013). Mobile majority: U.S. smartphone ownership tops 60%. Neilson Newswire. June 6. Neilson Company. Retrieved January 23, 2015. http://www.nielsen.com/us/en/newswire/2013/mobile-majority--u-s--smartphone-ownership-tops-60-.html.

O'Brien, E., Konrath, S. H., Grühn, D., & Hagen, A. L. (2013). Empathic concern and perspective taking: Linear and quadratic effects of age across the adult life span. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(2), 168–175.

Oliner, S. P., & Oliner, P. (1992). *Altruistic personality: Rescuers of Jews in Nazi Europe*. New York: Simon and Schuster.

Padilla-Walker, L. M., Coyne, S. M., & Fraser, A. M. (2012). Getting a high-speed family connection: Associations between family media use and family connection. *Family Relations*, *61*(3), 426–440.

Panek, E. T., Nardis, Y., & Konrath, S. (2013). Mirror or megaphone? How relationships between narcissism and social networking site use differ on Facebook and Twitter. *Computers in Human Behavior*, *29*(5), 2004–2012. doi:10.1016/j.chb.2013.04.012.

Park, L. G., Howie-Esquivel, J., & Dracup, K. (2014). A quantitative systematic review of the efficacy of mobile phone interventions to improve medication adherence. *Journal of Advanced Nursing*, *70*(9), 1932–1953.

Penner, L. (2002). Dispositional and organizational influences on sustained volunteerism: An interactionist perspective. *Journal of Social Issues*, *58*(3), 447–467.

Penner, L., Fritzsche, B. A., Craiger, J. P., & Freifeld, T. R. (1995). Measuring the prosocial personality. *Advances in Personality Assessment*, *10*, 147–163.

Penner, L., & Orom, H. (2009). Enduring goodness: A person by situation perspective on prosocial behavior. In M. Mikulincer & P. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature*. Washington, DC: American Psychological Assocation, Chapter 3. Retrieved from: https://www.med.wayne.edu/fam/faculty/pdfs/Penner%20and%20Orom%20Enduring%20Goodnessfinaledit_jan13.pdf

Perry, M. A. (1975). Modeling and instructions in training for counselor empathy. *Journal of Counseling Psychology*, *22*(3), 173.

Pettigrew, J. (2009). Text messaging and connectedness within close interpersonal relationships. *Marriage & Family Review*, *45*(6-8), 697–716.

Pew Research Center. (2012). Mobile Technology Fact Sheet. http://pewinternet.org/Commentary/2012/February/Pew-Internet-Mobile.aspx.

Postman, N. (1992). *Technopoly: The surrender of culture to technology*. New York: Vintage.

Przybylski, A. K., & Weinstein, N. (2013). Can you connect with me now? How the presence of mobile communication technology influences face-to-face conversation quality. *Journal of Social and Personal Relationships*, *30*(3), 237–246.

Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Robinson, J., Zahn-Waxler, C., & Emde, R. N. (2001). Relationship context as a moderator of sources of individual difference in empathic development. In Emde, R. & Hewitt, J. (Eds.) *Infancy to early childhood: Genetic and environmental influences on developmental change*, 257–268, New York, NY: Oxford University Press.

Rushton, P. (1980). *Altruism, socialization, and society*. Englewood Cliffs, NJ: Prentice-Hall.

Rushton, P. (1981). The altruistic personality. In P. Rushton & R. Sorrentino (Eds.), *Altruism and helping behavior*, chap.12. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rushton, P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*(1), 18.

Rushton, P., Chrisjohn, R. D., & Fekken, C. G. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, *2*(4), 293–302.

Rushton, P., Fulker, D. W., Neale, M. C., Nias, D. K., & Eysenck, H. J. (1986). Altruism and aggression: the heritability of individual differences. *Journal of Personality and Social Psychology*, *50*(6), 1192.

Rushton, P., Fulker, D. W., Neale, M. C., Nias, D. K., & Eysenck, H. J. (1989). Ageing and the relation of aggression, altruism and assertiveness scales to the Eysenck Personality Questionnaire. *Personality and Individual Differences*, *10*(2), 261–263.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*(5), 607.

Schwarz, N., & Bless, H. (1992a). Assimilation and contrast effects in attitude measurement: An inclusion/exclusion model. *Advances in Consumer Research*, *19*(1), 72–77.

Schwarz, N., & Bless, H. (1992b). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In Martin, L, & Tesser, A. (Eds). *The Construction of Social Judgments*, 217–245, Hillsdale, NJ: Lawrence Erlbaum.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195–202.

Shaw, R., & Bosworth, H. (2012). Short message service (SMS) text messaging as an intervention medium for weight loss: A literature review. *Health Informatics Journal*, *18*(4), 235–250.

Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, *27*(1), 3–23.

Sprecher, S., & Regan, P. C. (2002). Liking some things (in some people) more than others: Partner preferences in romantic relationships and friendships. *Journal of Social and Personal Relationships*, *19*(4), 463–481.

Staub, E. (1974). Helping a distressed person: Social, personality, and stimulus determinants. *Advances in Experimental Social Psychology*, *7*, 293–341.

Steffen, P., & Masters, K. (2005). Does compassion mediate the intrinsic religion-health relationship? *Annals of Behavioral Medicine*, *30*(3), 217–224. doi:10.1207/s15324796abm3003_6.

Stolinski, A. M., Ryan, C. S., Hausmann, L. R., & Wernli, M. A. (2004). Empathy, guilt, volunteer experiences, and intentions to continue volunteering among buddy volunteers in an AIDS organization. *Journal of Applied Biobehavioral Research*, *9*(1), 1–22.

Strayer, J. (1980). A naturalistic study of empathic behaviors and their relation to affective states and perspective-taking skills in preschool children. *Child Development*, *51*(3), 815–822.

Strayer, J., & Roberts, W. (2004). Empathy and observed anger and aggression in five-year-olds. *Social Development*, *13*(1), 1–13.

Taylor, S. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, *35*, 297–310.

Teherani, A., Hauer, K. E., & O'Sullivan, P. (2008). Can simulations measure empathy? Considerations on how to assess behavioral empathy via simulations. *Patient Education and Counseling*, *71*(2), 148–152.

Truax, C. B., Wargo, D. G., Frank, J. D., Imber, S. D., Battle, C. C., Hoehn-Saric, R., . . . Stone, A. R. (1966). Therapist empathy, genuineness, and warmth and patient therapeutic outcome. *Journal of Consulting Psychology*, *30*(5), 395.

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. New York: Basic.

Twenge, J., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, *92*(1), 56–66.

Twenge, J., & Foster, J. (2008). Mapping the scale of the narcissism epidemic: Increases in narcissism 2002–2007 within ethnic groups. *Journal of Research in Personality*, *42*(6), 1619–1622. doi:10.1016/j.jrp.2008.06.014.

Twenge, J., Konrath, S., Foster, J., Campbell, W. K., & Bushman, B. (2008). Egos inflating over time: A cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality*, *76*(4), 875–902; discussion 903–918. doi:JOPY507 [pii] 10.1111/j.1467-6494.2008.00507.x.

Underwood, B., & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin*, *91*(1), 143.

Unger, L. S., & Thumuluri, L. K. (1997). Trait empathy and continuous helping: The case of voluntarism. *Journal of Social Behavior & Personality*, *12*(3), 785–800.

Van der Mark, I. L., Ijzendoorn, M. H. v., & Bakermans-Kranenburg, M. J. (2002). Development of empathy in girls during the second year of life: Associations with parenting, attachment, and temperament. *Social Development*, *11*(4), 451–468. doi:10.1111/1467-9507.00210.

Van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A.-W., Lentjes, E. G., & Baron-Cohen, S. (2011). Testosterone administration impairs cognitive empathy in women depending on

second-to-fourth digit ratio. *Proceedings of the National Academy of Sciences*, *108*(8), 3448–3452.

Volbrecht, M. M., Lemery-Chalfant, K., Aksan, N., Zahn-Waxler, C., & Goldsmith, H. H. (2007). Examining the familial link between positive affect and empathy development in the second year. *The Journal of Genetic Psychology*, *168*(2), 105–130.

Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, *93*(5), 845.

Warner, R. E. (1984). Enhancing teacher affective sensitivity by a videotape program. *Journal of Educational Research*, 366–368.

Waxman, H. C. (1983). Effect of teachers' empathy on students' motivation. *Psychological Reports*, *53*(2), 489–490.

West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., & Shanafelt, T. D. (2006). Association of perceived medical errors with resident distress and empathy. *JAMA: The Journal of the American Medical Association*, *296*(9), 1071–1078.

Whittaker, R., Borland, R., Bullen, C., Lin, R. B., McRobbie, H., & Rodgers, A. (2009). Mobile phone-based interventions for smoking cessation. *Cochrane Database of Systematic Review*, *4*, Article No.: CD006611, doi: 10.1002/14651858.CD006611.pub2.

Wilhelm, M. O., & Bekkers, R. (2010). Helping behavior, dispositional empathic concern, and the principle of care. *Social Psychology Quarterly*, *73*(1), 11–32.

Wink, P., & Dillon, M. (2002). Spiritual Development across the Adult Life Course: Findings from a Longitudinal Study. *Journal of Adult Development*, *9*(1), 79–94. doi:10.1023/a:1013833419122.

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology*, *16*(1), 155–188.

Wood, A. M., Brown, G. D., Maltby, J., & Watkinson, P. (2012). How are personality judgments made? A cognitive model of reference group effects, personality scale responses, and behavioral reactions. *Journal of Personality*, *80*(5), 1275–1311.

Yakimovich, D., & Saltz, E. (1971). Helping behavior: The cry for help. *Psychonomic Science*, *23*(6), 427–428.

Young, S. K., Fox, N. A., & Zahn-Waxler, C. (1999). The relations between temperament and empathy in 2-year-olds. *Developmental Psychology*, *35*(5), 1189.

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The development of empathy in twins. *Developmental Psychology*, *28*(6), 1038.

Zhou, Q., Eisenberg, N., Losoya, S. H., Fabes, R. A., Reiser, M., Guthrie, I. K., . . . Shepard, S. A. (2002). The Relations of Parental Warmth and Positive Expressiveness to Children's Empathy-Related Responding and Social Functioning: A Longitudinal Study. *Child Development*, *73*(3), 893–915.

PART } VII

Character Development

20 }

The Emergence of Moral Character in Infancy: Developmental Changes and Individual Differences in Fairness Concerns and Prosocial Behavior during the First Two Years of Life

Jessica A. Sommerville

1. The Origins of Character: Moral Concerns and Behavior as Cornerstones of Character

The issue of character occupies a central role in scholarship in both philosophy and psychology. In its lay use, *character* refers to characteristics that make an individual unique. More specifically, however, philosophers and psychologists are often concerned with character in the moral sense. That is, the constellation of moral qualities—both psychological (e.g., possessing appropriate moral concerns) and behavioral (e.g., acting in morally appropriate ways)—that a given person possesses that makes that individual distinct from other individuals (Homiak, 2011).

Scholars have long debated the nature of morally excellent character; that is, what makes for a good and virtuous person? Of equal importance and interest, both historically and currently, are questions regarding the origins and source of character: when and how does character develop? Historically, perspectives have ranged from claims that children are born a blank slate (Locke, 1690/1970), to suggestions that children are either inherently good (Rousseau, 1762/1979) or naturally selfish and sinful (Freud, 1901/1966; Hobbes, 1651/1996). This chapter addresses the emergence of character by investigating the development of moral concerns and behavior in the first two years of life.

1.1 FAIRNESS: A CASE STUDY

I focus specifically on the emergence of fairness concerns in the course of development, and how these relate to infants' prosocial behavior. For adults, issues of fairness and justice are pressing moral concerns. As numerous studies have now shown, adults care about procedural justice (i.e., that the processes by which decisions are made are fair and just; Lind & Tyler, 1988), distributive justice (i.e., that goods and resources are distributed in a fair and just manner; Deutsch, 1975), and retributive justice (i.e., that wrongdoings are punished in fair and just ways; Darley & Pittman, 2003). Fairness concerns are prominent in situations ranging from everyday social interactions to workplace decisions and behavior, and extending to societal-level policies and legal decisions. Thus, fairness is an ideal case study for examining the development of moral concerns and behavior.

Much empirical work has focused on the principles that individuals use to guide their judgments of what is fair and just. With respect to distributive justice, numerous studies in experimental economics have used bargaining games to reveal the underlying norms and principles that guide adults' distributions of goods and resources (e.g., Fehr & Fischbacher, 2003; Henrich et al., 2005). These findings have revealed that even in situations in which an individual is dividing resources between her- or himself and an anonymous social partner, and even when there is no opportunity for retribution from the social partner, adults appear to abide by the "principle of equality" (Deutsch, 1975); all other things considered, goods should be divided equally to recipients. In addition to dividing goods roughly equally to themselves and social partners, adults punish individuals who do not act fairly and act to redistribute goods equally when norms of fairness have been violated (Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Johnson, Dawes, Fowler, McElreath, & Smirnov, 2009). Moreover, distributing goods equally between oneself and a recipient also appears to be intrinsically rewarding: value-related neural structures are activated when an individual allocates goods equally; whereas unequal allocations engage brain regions typically associated with negative or aversive emotional states, such as disgust and pain (e.g., the anterior insula; Zaki & Mitchell, 2011).

These findings, taken together, suggest that concerns about fairness are a central part of adults' moral concerns, and, more specifically, that in situations in which goods and resources can be distributed, in the absence of additional knowledge, adults advocate dividing resources equally as fair and just. A critical question, then, is how and when children develop a sense of fairness. Since fairness concerns are central to mature moral cognition and behavior in adults, identifying when and how children develop a sense of distributive fairness can shed light on when moral considerations emerge in development and how best to facilitate such considerations.

1.2 THE IMPORTANCE OF STUDYING DEVELOPMENT: TIMING IS EVERYTHING

I focus specifically on the emergence of distributive fairness. A central aim of the chapter is to describe and examine the extent to which early sensitivity to fairness conforms to and/or departs from more mature notions of fairness. To this end, the main themes that are addressed concern the nature of infants' fairness representations, whether individual differences exist in such representations, and how fairness is integrated with competing socio-moral concerns. A second aim of the chapter is to characterize the developmental trajectory of fairness concerns. Understanding how a sensitivity to fairness develops is critical to a host of scholarly and pragmatic issues. For example, identifying whether a sense of fairness is present from the get go, or develops over time, allows researchers to understand whether sensitivity to fairness is primarily influenced by innate genetic factors or whether such a sensitivity derives from experience. At a pragmatic level, understanding the extent to which experience influences the emergence of fairness concerns (and other socio-moral concerns), and what types of experience appear to matter, paves the way for facilitating and fostering a sense of fairness and positive character development more broadly.

2. The Development of Fairness Concerns

2.1 PRIOR PERSPECTIVES ON THE DEVELOPMENT OF FAIRNESS CONCERNS

Initial work examining whether children abide by the principle of equality in distributing resources was conducted using experimental economic games, similar to those used with adults. In one study, children played an allocation game in which they could allocate candy to themselves and another person (Fehr, Bernhard, & Rockenbach, 2008). Across a series of conditions, children could select between egalitarian allotments (1:1) or allotments that created either advantageous inequality (e.g., 1:0, 2:1) or disadvantageous inequality (1:2) for the self. Across all conditions there was a significant increase in egalitarian choices from ages 3–4 to ages 7–8, and it was not until ages 7 or 8 that the majority of children systematically preferred the egalitarian options, suggesting that the tendency to distribute goods fairly and equally awaits mid-childhood.

However, these initial studies had several limitations, presenting a challenge for interpreting age-related changes in children's task performance. Specifically, tasks used with children often featured inhibitory control demands, requiring children to strongly act against their own interests. These tasks also had low ecological validity; children were often asked to act toward an absent and anonymous social partner. Thus, it is possible that young children's

abilities were underestimated. Indeed, additional research is consistent with this claim. In more ecologically valid contexts, preschool-age children provide evidence of an awareness of, and adherence to, equality: 3.5-year-old children distribute toys equally to recipients in a third-party task (Olson & Spelke, 2008), verbally protest unfair distributions in naturalistic settings (particularly when they are disadvantaged; LoBue, Nishida, Chiong, DeLoache, & Haidt, 2011), and share the fruits of their labor equally with a social partner following completion of a collaborative task (Warneken, Lohse, Melis, & Tomasello, 2011). These findings suggest that a basic sense of fairness is present by at least the early preschool years.

2.2 INVESTIGATING FAIRNESS CONCERNS IN INFANCY: TESTING FOR THE PRESENCE OF FAIRNESS EXPECTATIONS.

Our goal was to determine whether a basic sense of fairness could be traced back even further in life, to infancy. Although classic developmental theories posited that cognitive capacities did not emerge until the end of the second year of life (e.g., Piaget, 1953), contemporary work has overturned this notion. Infants understand basic principles underlying the physical (Spelke, 1994) and social world (Woodward, 2013), possess at least implicit forms of self-awareness (Rochat, 2003), produce empathic responses (Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992), and can even engage in some forms of abstract reasoning (Marcus, Vijayan, Rao, & Vishton, 1999). Most importantly for our interests, recent work suggests that prosocial and additional other-regarding preferences appear to flourish between 14 and 18 months of age (Warneken & Tomasello, 2006, 2007). Thus, we started our research by asking whether 15-month-old infants possessed a basic sense of fairness.

2.2.1 An Initial Study

Our specific question of interest was whether 15-month-old infants were able to detect violations to fairness norms in the context of a third-party distribution task (i.e., in a task in which infants watched an individual distribute goods to recipients). To address this question we created a novel paradigm that we labeled the *violation-of-expectancy resource distribution task* (henceforth, RDT). This paradigm relied on the firmly established finding that infants tend to show enhanced visual attention (i.e., they tend to look longer) to events that are unusual, unexpected, or that violate a preexisting expectation than to events that are typical, expected, or consistent with a preexisting expectation (Baillargeon, Wasserman, & Spelke, 1985). In these paradigms, infants are typically shown the first part of an event during an initial familiarization phase; the outcome is initially occluded or not depicted. On test events infants see outcomes that complete the event in ways that are consistent with, or violate,

adult conceptual notions. If infants show enhanced visual attention (i.e., if they look longer) to events that violate adults' conceptual notions, then a nascent sensitivity to those concepts is inferred. Control conditions are used to rule out competing interpretations, such as the possibility that enhanced looking to the conceptually violating event is driven by low-level perceptual processes.

During the RDT, infants first watched a distribution movie in which an adult actor (the distributor) sat at a table with two recipients. At the request of the recipients, the distributor proceeded to allocate a bowl of crackers (in one movie) or a pitcher of milk (in a second movie) to the recipients. During the allocation a black occluding screen covered the actors' plates/glasses and the contents of the bowl or pitcher so that the exact amount distributed to each recipient was hidden. On test trials, infants saw static outcomes in which the results of the distributive actions were revealed: during the fair outcome, both recipients received the same amount of crackers (or milk); during the unfair outcomes, one actor received more crackers (or milk) than the other. Control trials showed only the physical outcomes (e.g., the crackers on the plate or milk in the glasses), without the actors present, to ensure that the infants did not have a preexisting preference for symmetrical or asymmetrical displays. All infants saw both experimental and control trials for each movie.

Our prediction was that if the infants were aware of norms of fairness, they should show enhanced visual attention to the unfair over the fair outcome. Specifically, they should look longer to the outcomes that featured an unequal allotment of crackers versus an equal allotment of crackers. Critically, we also predicted that infants' looking times to the outcomes featured in the control trials would not differ. Together, this pattern of findings would suggest that infants' longer looking to the unfair outcomes on the experimental trials was driven by concerns about fairness per se. Equivalent looking to these same outcomes in the control trials (in which recipients were absent) would rule out the possibility that the findings in the experimental condition were merely driven by a low-level perceptual bias (i.e., rule out the possibility that infants merely prefer to look at asymmetrical versus symmetrical outcomes).

Infants also took part in a novel prosocial task that we created called the *altruistic sharing task*. Our motivation for including this task was twofold. First, we sought to validate the RDT using a second, ancillary task. Second, we sought to investigate potential linkages between infants' fairness concerns and their prosocial behavior given theoretical speculation regarding the connections between fairness and altruism (Boehm, 1999; Fehr & Fischbacher, 2003), and in order to contribute to the debate regarding whether moral cognition and moral behavior are tightly interwoven or independent of one another (e.g., Blasi, 1980).

During the altruistic sharing task, infants were first given the opportunity to select between two toys, and their toy preference was noted. Then, prior to the test phase, the experimenter handed the infants the other toy, so that

they were now holding both their preferred toy and their non-preferred toy. During the test phase, an unfamiliar experimenter approached the infants and produced an ambiguous request ("Can I have one?") while looking directly at the infant, to avoid cuing the infant, and holding out her hands. This experimenter was unaware of the study hypotheses. We recorded whether the infant shared a toy, and if so, which of the two toys that he or she shared. In advance of prior work, our novel task allowed us not only to assess whether infants engaged in prosocial behavior (which is affected by a myriad of factors, including infants' understanding of the experimenter's desire, infants' understanding that the experimenter is producing a request, infants' understanding that the request is directed at them, etc.), but also to specifically identify the nature of infants' prosocial motivation. For infants who shared one of the two toys, we could ensure that they understood the task directions and did not suffer from stranger anxiety; thus, examining which of the two toys the infants chose to share allowed us to identify infants that shared more altruistically, by sharing their preferred toy, versus those that shared more selfishly, by sharing their non-preferred toy.

The results from the RDT indicated that the infants looked significantly longer at the unfair outcome over the fair outcome, suggesting that by 15 months of age, infants possess expectations about how goods are typically distributed and, specifically, expect goods to be distributed equally to recipients. Critically, infants' looking times to the control outcomes did not differ, suggesting that the results from the experimental condition were not due to a general preference for asymmetrical over symmetrical outcomes. This pattern was found for the distribution of both the cookies and the milk. Thus, by 15 months of age, infants appear to expect resources to be distributed fairly in a resource distribution task (see Geraci and Surian, 2011; and Sloane et al., 2012, for similar results with slightly older infants). Given the centrality of fairness to moral concerns more broadly, these findings also provide evidence for emerging moral sensitivity in the second year of life.

Our findings also revealed variability in infants' performance on the RDT. This variability was systematically related to their performance on the altruistic sharing task. Infants who shared their preferred toy looked longer to the unfair outcome than the fair outcome; whereas infants who shared their non-preferred toy showed the reverse pattern of findings. Critically, performance on the sharing task was unrelated to infants' attention to the control outcomes. These additional findings help to interpret the results of the RDT. Specifically, our results suggest that it is unlikely that infants' enhanced attention to the unfair outcome merely results from statistical learning (i.e., infants' detection of observable regularities in the world). If this were the case, there would be no reason to expect a tight relation between infants' prosocial behavior and their performance on the RDT. More broadly, our results provide evidence that there are individual differences in infants' fairness concerns at this

age that relate to their willingness to generously share toys. These findings give teeth to the theoretical notions that fairness and altruism are intertwined and that socio-moral judgments or evaluations and behavior are closely interconnected.

As a whole, our results (Schmidt & Sommerville, 2011; Sommerville et al., 2013) provide some of the first evidence that by the second year of life infants demonstrate an awareness of fairness norms, in that they expect resources to be distributed equally to recipients. Critically, like older children and adults there is individual variability in infants' fairness concerns. These important findings opened the door for us to ask critical questions about the developmental origins of infants' fairness concerns, as well as the sources and stability of individual differences in such concerns.

2.2.2 Tracing Developmental Changes in Infants' Fairness Concerns

The next question we asked was whether fairness concerns are continuous across the course of infancy or whether infants' sensitivity to fairness undergoes developmental change. Addressing questions of continuity versus discontinuity in development is critical to informing debates in moral and social cognitive development. For example, one such debate concerns whether socio-moral concerns or sensitivities are innate, or constructed from experience. More broadly, findings regarding the developmental trajectory of fairness concerns and similar moral concerns speak to debates regarding the origins of character. Specifically, investigating the developmental trajectory of fairness concerns and other moral sensitivities can help to reveal the relative contributions of genetic and experiential factors to the development of character.

Thus, our next question was whether infants' fairness concerns were present and continuous across infancy, or whether such concerns show age-related changes. If fairness concerns are present across infancy, then we can conclude that such concerns are likely innate. In contrast, if such concerns emerge in the course of development, these findings would suggest an important role for experience in the development of fairness concerns.

To address this question, we conducted a cross-sectional study with infants at 6, 9, 12, and 15 months of age ($n = 32$ per age group; Ziv & Sommerville, under review). Infants received a variant of our RDT paradigm that was similar to prior studies. Specifically, infants watched an occluded distribution of crackers that resulted in either a fair or unfair outcome, along with a matched control condition, in which a distributor distributed crackers to inanimate "recipients" that consisted of two pillows. The modification of the control condition from prior studies allowed us to rule out the possibility that infants' longer looking to unfair outcomes results from a more general expectation about how goods are typically sorted (i.e., into equal vs. unequal piles). In addition, infants at 12 months and 15 months of age received our altruistic sharing task. Because a pilot sample of 9-month-olds that were tested in the altruistic sharing task

provided no evidence of sharing behavior during lab-based tasks, we interviewed parents of 9-month-old infants about their naturalistic sharing behavior. Specifically, we asked about the onset of infants' sharing behavior—that is, if infants were sharing at all, and if so, whether they were sharing with only their primary caregiver or also with other people (including siblings and other adults). Pilot work revealed that no parents of 6-month-old infants reported sharing behavior in any context. Thus, six-month-old infants completed only the RDT.

Our study was designed to test two key hypotheses: first, we hypothesized that there would be a developmental transition in infants' fairness concerns as measured via the RDT. This pattern of findings would be consistent with the idea that fairness concerns emerge from infants' real-world experience (vs. being innate). Second, we hypothesized that this transition would coincide with the onset of the infants' sharing behavior in naturalistic settings as captured by our parental interview. This second hypothesis was based on our earlier findings demonstrating linkages between how older infants share toys and their fairness concerns. It is also based on our belief that the opportunity to take part in sharing exchanges allows infants to experience being both the agent and recipient of fair behavior and unfair behavior, and thus provides them with the opportunity to learn about values like equality and reciprocity. We hypothesized that these experiences would be critical to infants' developing notions of fairness. Finally, we sought to replicate the relation between infants' fairness sensitivity and their altruistic sharing behavior that we had demonstrated in prior work (e.g., Schmidt & Sommerville, 2011; Sommerville et al., 2013).

Our results revealed a clear developmental shift in the infants' fairness concerns. Infants at 12 months and 15 months of age looked significantly longer to the unfair (vs. fair) outcome, replicating previous findings. Infants at 6 months of age looked equally to the unfair and fair outcomes, providing no evidence of sensitivity to fairness. Critically, infants at 9 months of age, as a group, were transitional in their fairness concerns: infants looked marginally longer at the unfair outcome (vs. the fair outcome). Variability within the sample at this age was related to parental reports of sharing status. Infants whose parents reported no sharing behavior did not differ in their looking to the outcomes. Infants whose parents reported selective sharing with the primary caregiver looked marginally longer to the unfair outcome versus the fair outcome. Finally, infants whose parents reported sharing with the primary caregiver as well as sharing more broadly looked significantly longer at the unfair outcome versus the fair outcome. At all ages, infants looked equally at the unequal and equal events (when pillows replaced people as participants), suggesting that longer looking in the experimental condition cannot be attributed to perceptual preferences for unequal outcomes or to expectations about how goods should be sorted or categorized.

Taken together, these findings provide evidence of a developmental shift in infants' fairness concerns at 9 months of age. Moreover, our results suggest that the onset of sharing behavior is associated with sensitivity to fairness-norm violations. These findings raise the possibility that sharing exchanges may provide a means by which parents transmit values that are central to fairness, such as reciprocity and equality. We believe that sharing exchanges, given their turn-taking structure and emphasis on reciprocity, may present powerful opportunities to learn about fairness. In ongoing work we are testing this hypothesis by investigating the impact of sharing interventions on younger infants' fairness concerns. If exposing younger infants (e.g., 7- to 8-month-old infants) to sharing interventions increases their sharing behavior and leads to a sensitivity to fairness norms, these findings would provide support for the claim that exposure to and participation in sharing interactions starting at about 9 months of age helps infants to establish a basic sense of fairness.

2.3 THE NATURE OF FAIRNESS CONCERNS IN INFANCY

The findings reviewed so far are consistent with the possibility that infants possess at least an implicit awareness of fairness norms. The fact that individual variability in infants' fairness concerns aligns with their prosocial behavior suggest that it is unlikely that infants' performance on the RDT occurs merely because infants are good at picking up the statistical regularities that are present in the world. For example, it is unlikely that infants have merely learned through visual experience that distributions of goods typically result in equal allotments of goods or resources.

The next question we investigated was whether infants go beyond merely detecting fairness norms, to evaluating actions and individuals on the basis of such norms. Adults not only notice when fairness norms have been violated, but also evaluate both distributive actions, and the individuals that produce such actions, by assigning positive value to fair behavior and negative value to unfair behavior. For example, adults engage in altruistic punishment, by incurring a cost to the self, in order to correct inequality through the redistribution of goods and resources (Fehr & Gächter, 2002).

2.3.1 Does Sensitivity to Fairness Affect Infants' Affiliative Behavior?

To begin to unpack the underlying nature of infants' fairness concerns, Monica Burns and I (Burns & Sommerville, 2014) created a novel affiliative choice paradigm to investigate whether 15-month-old infants select social partners on the basis of prior fair versus unfair behavior. In advance of our prior work, demonstrating that infants show enhanced visual attention to unfair (vs. fair) outcomes, the current study allowed us to ask whether infants' sensitivity to violations to fairness norms impacted their social behavior. Evidence for the systematic selection of fair over unfair individuals would demonstrate that

infants were engaging in (at least primitive) evaluations of fair and unfair individuals, and, more specifically that infants more positively evaluated fair individuals than unfair individuals.

Infants first watched a live distribution phase in which one actor distributed toys equally to two recipients on two different occasions, and another actor distributed toys unequally to two recipients on two different occasions, consistently advantaging one recipient over another recipient. Our question of interest in this study was whether infants would use the actor's prior history of fair or unfair behavior to guide their affiliative behavior during the subsequent choice phase.

The choice phase consisted of three test trials. On each trial, infants were given the opportunity to affiliate with either the (previously) fair or (previously) unfair individual. During the first and second test trial, the two actors sat at a table across from the infant and simultaneously offered infants identical toys. We coded which actor the infants selected a toy from. On the third test trial, the actors sat at opposing sides of the room stacking blocks in the same manner; infants were released by parents at a point exactly midway between the actors. We coded the actor whom the infant approached to play with.

Critically, on each test trial the actors' behavior was equated and the primary experimenter confirmed that the actors acted identically during the test trial. Moreover, the identity of the actor playing the fair distributor was counterbalanced across infants (deconflating personal identity and fair/unfair behavior), and the distributors were actually made unaware of whether they were distributing toys fairly or unfairly (see Burns & Sommerville, 2014, for details); thus, any systematic selections of a given actor must be based on the actor's prior fair or unfair behavior, per se, rather than any other factors.

The infants consistently selected the fair actor more frequently than the unfair actor, and at rates above chance, during the test trials. Thus, these findings demonstrate that above and beyond the sheer ability to detect violations of fairness norms, infants also use an actor's social history of acting fairly or unfairly to guide their social selections and affiliative behavior. Taken together, these findings suggest that infants' reactions to fair and unfair behavior, and fair and unfair outcomes, are not merely reactions to statistically expected versus anomalous events. Rather, infants also appear to ascribe meaning or value to these events: infants (more) positively evaluate fair individuals over unfair individuals. These findings contribute strongly to our understanding of the nature of infants' fairness concerns and their reactions to moral events more broadly.

2.4 INDIVIDUAL DIFFERENCES IN FAIRNESS CONCERNS IN INFANCY

Although past work has revealed that adults, as a group, endorse a fair and equal distribution of goods to recipients, there are of course individual and cultural differences in how closely adults subscribe to and adhere to fairness

norms (Henrich, Heine & Norenzayan, 2010; Hsu, Anen, & Quartz, 2008). One critical question, then, concerns when these individual differences emerge in development.

The results of our initial study (Schmidt & Sommerville, 2011) provide evidence for individual variability in performance linked to infants' prosocial behavior. Of course, one critical question concerns whether such individual variability reflects stable dispositional individual differences. Given the difference in surface features in the RDT versus the altruistic sharing task, our initial findings lend support to the idea that variability in infants' performance may reflect dispositional individual differences. However, given that the infants took part in both tasks during the same testing session, an alternate possibility is that a subset of infants in the sample may merely have been particularly "prosocially motivated" on that day; that is, the individual variability that we have witnessed in infants' performance across both tasks may be transient not stable.

2.4.1 Relations between Infants' Fairness Concerns and Parental Empathy

To begin to address both the questions regarding the nature of individual variability and the source of potential individual differences, we first asked whether variability in infants' fairness concerns were related to parental dispositional characteristics. If parental dispositional characteristics predict individual variability in infants' fairness concerns these findings would provide support for the notion that individual variability in our tasks reflects actual individual differences. In addition, associations between these measures would position us to begin to identify the *source* of individual differences in fairness concerns in infancy.

Scholars have suggested that people are motivated to act in a fair and just manner based on their understanding of the emotional consequences of being treated unfairly (Hoffman, 2001)—that is, based on their ability to experience empathy toward those who are the victims of unfair treatment. Thus, in our first attempt to begin to identify the source of individual variability in infants' fairness concerns we asked whether individual differences in parental empathy might predict variability in infants' fairness concerns.

We investigated links between parental empathy and infants' fairness concerns in two separate studies. In our previously mentioned cross-sectional study (Ziv & Sommerville, under review), parents of 12- and 15-month-old infants completed the Davis Interpersonal Reactivity Inventory (IRI; Davis, 1980, 1983), a self-report measure of dispositional empathy. This measure is designed to assess both the cognitive components of empathy (the ability to take someone else's perspective and appreciate the significance of another person's perspective), and the affective components of empathy (feelings of concern in the face of another person's distress). The measure contains four subscales: perspective taking (cognitive empathy), empathic concern (affective

empathy), fantasy (the tendency to readily imagine oneself in fictional situations), and personal distress (the extent to which individuals feel personally uncomfortable or distressed by another person's suffering). We predicted a selective association between parental scores on the cognitive and emotional empathy subscales and infants' fairness sensitivity (operationalized as their attention to the unfair outcome over the fair outcome in the RDT). In another study, we investigated whether parental performance on the Davis IRI would predict infants' tendency to select the fair actor over the unfair actor in the affiliative choice paradigm, which assessed how consistently infants chose the fair actor to interact with.

Across both studies, we found a significant positive association between parental self-reported dispositional empathy and infant fairness concerns: parental empathy scores predicted the extent of infants' selective attention to the unfair outcome in the RDT and infants' frequency of selection of the fair actor in the affiliative choice paradigm (Sommerville et al., in preparation). Although both the cognitive and affective components of empathy related to infants' fairness concerns, there was evidence that infants' fairness concerns were more tightly and consistently linked to parental perspective-taking scores. One possibility, then, is that fairness concerns are more tightly interrelated with empathic perspective taking (e.g., the consideration and balancing of the welfare of both recipients) than with empathic concern (e.g., concern about the diminished welfare of the disadvantaged recipient).

2.4.2 Understanding the Nature of the Association between Infant Fairness Concerns and Parental Empathy

One possible explanation for the association between infant fairness concerns and parental empathy is that parents who are more empathic are more concerned about issues of equality and reciprocity, and thus emphasize these values in their interactions with infants. For example, parents who are higher in empathy may vary from parents lower in empathy in the frequency with which they encourage infants to share (and in the frequency with which they share with infants) and in the nature of sharing exchanges (e.g., the extent to which they encourage more equality-based, reciprocal exchanges). This may account for our findings that both the emergence of sharing behavior is linked to the onset of fairness concerns (at the group level) at 9 months of age, and that individual differences in the way that infants share toys predict individual differences in infants' fairness concerns at 12 months to 15 months of age. In ongoing work, we are examining whether interventions designed to enhance empathy in parents have down-stream effects for infants. Our hypothesis is that interventions that improve parental empathic tendencies will lead to changes in the frequency and nature of parent-infant sharing interactions and ultimately facilitate infant fairness concerns.

Another possible reason for the association between parental self-reported empathy and infants' fairness concerns is that the association between parental empathy and infants' fairness concerns reflects a shared genetic tendency toward prosociality or other-regarding preferences. Of course, it is likely that these parent-infant relations emerge from a joint combination of and interaction between genetic and experiential factors.

While finding an association between parental empathy and infant fairness concerns may seem intuitive at some level, it is notable that we were able to capture this association. This relation was demonstrated across two different individuals (parent and infant), using three different tasks with completely different surface features (one using infant looking times to resource distribution outcomes, one relying on infants' consistency of selecting the fair actor to engage with, and the other relying on parent self-reported questionnaire data). Moreover these measures were completed at two different points in time (parents completed the questionnaire prior to their lab visit, and thus prior to their infants' participation in the RDT and affiliative choice paradigm). These findings are consistent with the possibility that our infant tasks are capturing something enduring and dispositional about infant tendencies. In ongoing work, we are investigating the stability of early infant fairness concerns and prosocial behavior using longitudinal designs in which we test infants across time at closely spaced intervals to determine whether infants' performance on our fairness and sharing tasks is consistent across time.

2.5 INTEGRATING FAIRNESS CONCERNS WITH COMPETING SOCIAL AND MORAL MOTIVATIONS

In the real world, multiple social and moral concerns compete to guide our social behavior and our social evaluations. Consider the following example: after your weekly soccer game you are walking down the stairs behind your teammate and another player on the opposing team. Midway down the stairs both players encounter a wet step and simultaneously slip and fall. If the only concern motivating your social behavior were your concern about the general welfare of others, then you would be equally likely to help either player (or at least who you choose to help first would be random). However, it is likely that you would first come to the aid of your fallen teammate. This is because concerns about social category membership and moral concerns (among other motivations) are integrated in the service of guiding our social behavior and social evaluations (e.g., Wright, Dinsmore, & Kellaris, 2013). As this example nicely illustrates, in the real world, it is rare that our social behavior is guided by single or unitary motivations. Indeed, how we weigh or prioritize social concerns that are more self- or group serving against those that are more morally virtuous is central to the issue of character.

2.5.1 Do Infants Integrate Fairness Concerns with Other Social and Moral Concerns?

Do infants possess, and do they integrate, various social and moral concerns? Using our affiliative choice paradigm, Burns and I (Burns & Sommerville, 2014) investigated whether infants' social selections would change when the fairness status of an individual competed with other social information. In particular, we were interested in whether the race of the actors involved in the distribution events would impact infants' social selections. A variety of evidence suggests that infants show same-race preferences in their attention to faces: for example, infants as young as 3 months of age prefer to look at same-race over other-race faces (Kelly et al., 2005). Moreover, initial evidence from our laboratory suggested that such preferences influence infants' social selections as well: after watching Caucasian and Asian individuals engage in a neutral behavior (stacking blocks), Caucasian infants preferentially selected Caucasian (over Asian) individuals at rates above chance, suggesting a baseline preference for same-race individuals (Shin, Burns, Yun & Sommerville, 2011). Thus, our question was whether infants' preferences for fair or unfair individuals was also affected by the race of the distributors and recipients involved in the exchange.

To address this question, we tested eighty 15-month-old Caucasian infants across two experiments (we focused selectively on Caucasian infants to allow us to ask questions about infants' evaluations of same-race vs. other-race members; Burns & Sommerville, 2014). In the first experiment, we contrasted the infants' affiliative preferences in the baseline condition discussed in the section 2.3.1 (in which both distributors were Caucasian; henceforth, the "Caucasian fair/Caucasian unfair" condition) with another condition in which the Asian distributor was fair and the Caucasian distributor was unfair (henceforth, the "Asian fair/Caucasian unfair" condition). The distribution events were otherwise identical across conditions. The choice trials, in which infants had the opportunity to select between the (previously) fair and (previously) unfair individuals across three different trial types, were also matched across conditions.

The question addressed in the first experiment was whether the infants' preference for the fair actor would be attenuated under conditions in which the actor's social history (e.g., whether she had previously distributed toys fairly or unfairly) conflicted with the distributor's social group. Specifically, we wanted to know whether the infants' systematic preference for the fair actor that we found in the Caucasian fair/Caucasian unfair condition was attenuated in the Asian fair/Caucasian unfair condition. We found that infants in the Asian fair/Caucasian unfair condition showed a numerical preference for the fair actor, but that this preference was weakened in comparison to the Caucasian fair/ Caucasian unfair condition and did not differ significantly from chance performance. These results suggest that infants take into consideration

both an actor's prosocial history, and their social group membership, when making their social selections, and that when these factors conflict their preference for fair individuals is weakened.

Experiment 2 was designed to determine how infants make social selections under conditions in which an individual's history of fair versus unfair behavior is fully crossed with both the race of the distributors and recipients. One potential objection to our findings from the first experiment is that because infants' selection of the fair actor in the Asian fair/Caucasian unfair condition was not above chance, it may merely be the case that infants have difficulty making systematic social selections when multiple dimensions must be integrated. Thus in Experiment 2, across all conditions, infants saw an Asian distributor and a Caucasian distributor, as well as a Caucasian recipient and an Asian recipient. For half of the infants, the Caucasian distributor was fair, and the Asian distributor was unfair; for the remaining half, the Asian distributor was fair, and the Caucasian distributor was unfair. Also, for half of the infants, the unfair distributor gave more toys to the Caucasian recipient; for the remaining half of infants, the unfair distributor gave more toys to the Asian recipient. Thus, infants could make their social selections on the basis of one of three factors—they could systematically select the fair actor; they could systematically select the Caucasian actor; or they could systematically select the actor that advantaged same-race individuals (e.g., the actor that gave more to the Caucasian recipient). Alternately, they could base their choice on some combination of these factors.

The results revealed that infants' likelihood of selecting the fair actor was influenced by the identity of the recipient advantaged by the unfair actor. Specifically, when the unfair actor gave more toys to the Asian recipient than the Caucasian recipient infants were more likely to select the fair actor than when the unfair actor gave more toys to the Caucasian recipient than the Asian recipient. Another way of putting this is that infants were less concerned about inequality when a same-race member was the beneficiary of inequality than when an other-race member was a beneficiary of inequality. Importantly, our results provide evidence for integration of multiple dimensions in their affiliative behavior. If infants failed to integrate these competing dimensions then their actor selections would be uninfluenced by any of the experimentally manipulated variables and their affiliative choices would have been at chance levels across all conditions. Our findings suggest that infants take into account not only whether an individual behaves fairly or unfairly when selecting their social partners, but also the consequences of distributive actions for same-race and other-race members. Infants then, like adults, integrate competing social and moral concerns and use this information to guide their affiliative behavior.

Over and above demonstrating that infants integrate various social and moral concerns and use this information to guide their affiliative behavior, these findings provide new information regarding infants' ability to form and

utilize social categories, and the information that infants use to form such categories. Prior work has shown that infants possess preferences for individuals that share characteristics of the social group to which the infant belongs: for example, infants prefer to accept toys from native language speakers over non-native language speakers (Kinzler, Dupoux, & Spelke, 2007); preferentially imitate native language speakers over non-native speakers (Buttelmann, Zmyz, Daum, & Carpenter, 2013); and, more broadly, prefer others that are similar to the self, even along fairly superficial dimensions (Mahajan & Wynn, 2012). Past research has also yielded evidence for perceptual preferences for same-race members over other-race members (Kelly et al., 2005), but the evidence for social preferences for same-race individual has been equivocal (e.g., Kinzler & Spelke, 2011; Shin et al., 2011).

Our findings contribute to this literature in two ways. First, they suggest that race *does* affect infants' social preferences: infants are less likely to pick a fair distributor over an unfair distributor when the fair distributor is an other-race member (and the unfair distributor is a same-race member). Second, our findings suggest that infants may be beginning to make basic in-group/out-group distinctions, since they consider not only the race of the distributor in their affiliative decisions, but also the consequences of the distributive actions for same-race (in-group) members or other-race (out-group) members. Research has revealed that it is by the preschool years that children begin to form implicit positive evaluations of in-group members and implicit negative evaluations of out-group members based not only on race but also a range of different characteristics (Cvencek, Greenwald, & Meltzoff, 2011; Dunham, Chen, & Banaji, 2013); our findings do not necessarily indicate that infants negatively evaluate out-group members but, rather, are consistent with the possibility that infants are beginning to make basic in-group/out-group distinctions, and that race is one of the dimensions that infants use to make this distinction.

3. Back to Character: It Starts in Infancy

Our findings suggest that infants' emerging socio-moral concerns may form the basis for the development of character. Like adults, infants are sensitive to events and outcomes that are pertinent to the welfare of others: starting in the second year of life, if not earlier, infants recognize when norms regarding the fair and equal treatment of individuals have been violated and select social partners on the basis of their tendency to act fairly.

3.1 INFANT FAIRNESS CONCERNS AND MORAL SENSITIVITIES

Our results suggest that infants are sensitive to fairness in a particular context: they expect goods and resources to be distributed equally to recipients.

Of course, older children and adults are not limited to a single model of fairness. For example, we often consider that issues of merit (e.g., the quantity and quality of work as it relates to pay), and issues of need (e.g., who is more needy) when evaluating what is a fair and just outcome. Moreover, for older children and adults concerns of fairness are not limited to the distribution of goods or resources. As indicated earlier, we also care about procedural justice and retributive justice. Our findings, then, indicate that infants are in possession of one of the most basic forms of fairness; the questions of whether infants are also sensitive to other models of fairness and whether infants can apply fairness concerns in other domains await further inquiry. However, the fact that this basic form of fairness is acquired early on implies that the seeds of socio-moral concerns are in place early on, and are influenced by everyday experiences that do not rely on rich verbal abilities.

Our findings are also consistent with other work suggesting that by the end of the second year of life infants possess a range of other-regarding tendencies, extending to helping (Liszkowski, 2005; Warneken & Tomasello, 2006; 2007) and comforting behaviors (Svetlova, Nichols, & Brownell, 2010; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992). Our results are also consistent with research indicating that infants distinguish between agents on the basis of whether they help or hinder others (Hamlin, Wynn, & Bloom, 2007; Hamlin & Wynn, 2011; Hamlin, Wynn, Bloom, & Mahajan, 2011). Together, these results suggest that the basic building blocks of socio-moral concerns can be traced back to infancy. Our findings also add to this work by showing that individual differences in these concerns can also be found in infancy.

3.2 POTENTIAL IMPLICATIONS FOR MECHANISMS OF DEVELOPMENTAL CHANGE AND SOURCES OF INDIVIDUAL DIFFERENCES

Critically, above and beyond providing evidence for the origins of moral character in infancy, our work has the potential to yield important information about mechanisms of developmental change and sources of individual differences in infants' fairness concerns and prosocial tendencies. Specifically, we have found that the developmental onset of fairness concerns is empirically linked to the emergence of infants' ability to share toys with caregivers in naturalistic settings. Moreover, individual differences in infants' fairness concerns starting in the second year of life are linked to parental self-reported empathic tendencies. These findings suggest that infants' fairness concerns, and perhaps by extension other socio-moral concerns, are experientially influenced.

Although there is much to learn about the relations between sharing and fairness and between parental empathy and fairness, we can speculate regarding the underlying reasons for these associations. We believe that

individual differences in parental empathy contribute to both the emergence of infants' fairness concerns and to their individual differences in such concerns because parents with strong empathic tendencies may value reciprocity and equality, and transmit these values to their infants in the context of naturalistic interactions. We further speculate that sharing interactions, in particular, may be a particularly rich venue for the transmissions of such values, because such interactions are naturally characterized by reciprocity and equality. Consequently, we would predict that variability in the frequency with which parents and infants take part in sharing interactions and the nature of such interactions play a causal role in the development of fairness concerns and individual differences in such concerns.

Of course, these predictions are speculative at present; their confirmation or dismissal will require future empirical studies. We are currently directly testing these hypothesis through sharing intervention studies with pre-sharing infants (to determine whether exposure to sharing interventions facilitates the onset of sharing behavior and leads to the earlier acquisition of the awareness of fairness norms), and through intervention studies with adults aimed at increasing parental empathy (to determine whether changes in parental empathy have consequences for infants' fairness concerns and prosocial behavior). Critically, if our hypotheses are correct, the results of these studies will have important applied value. Specifically, they can be used as bases for interventions aimed at facilitating moral concerns and the development of character in early childhood more broadly.

3.3 CHARACTEROLOGICAL LIMITATIONS IN INFANCY

Our findings to date have also yielded important new information regarding characterological limitations in infancy. Specifically, our results suggest that infants are not *exclusively* prosocial. Indeed, we found that fairness concerns compete with other self- or group-motivated concerns: infants generally prefer fair individuals over unfair individuals; yet this preference is muted in circumstances in which members of the infants' own group may in fact benefit from the inequality. Again, like adults, infants appear to be concerned about what is good or right, but they are not concerned only with what is good or right generally; they are also concerned with what may be good for them or their group. Future work can establish the circumstances in which some concerns are activated over others, as well as the sources of individual differences in how infants integrate and weigh competing concerns.

3.4 CONCLUSIONS

Questions regarding the origins of character are central to not only psychology and philosophy, but also are at the heart of broader scholastic concerns

regarding what it means to be human. For centuries, scholars have debated when and how character emerges, as well as the nature of the "starting state" of moral concerns. Our findings suggest that, despite their differing perspectives, Locke, Rousseau, Freud, and Hobbes all got something right. As Rousseau would suggest, young children are inherently prosocial: they care about the welfare of others from early on. As Locke would suggest, experience matters: children's fairness concerns and prosociality are experientially influenced. As Hobbes and Freud would suggest, children can be selfish: children also possess concerns that are considerably more self-oriented or self-motivated. Taken together, our findings suggest that a key to character development, then, may be to identify and harness those experiences and circumstances that lead to increased consideration of the general welfare of others while balancing and integrating competing group- and self-motivated concerns.[1]

References

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, *88*(1), 1.

Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.

Burns, M. P., & Sommerville, J. A. (2014). "I pick you": The impact of fairness and race on infants' selection of social partners. *Frontiers in Psychology*, *5*, 93.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, *84*(2), 422–428.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The preschool implicit association test. *Journal of Experimental Child Psychology*, *109*(2), 187–200.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*(4), 324–336.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

[1]I would like to thank the parents and infants who generously donated their time to participate in this research. I am grateful to Monica Burns, Marco Schmidt, and Kayla Upshaw for their contributions to this work, as well to all the students and staff at the Early Childhood Cognition Lab at the University of Washington for their help with data collection and coding. I am grateful to the Early Childhood Cognition Lab, the Social Cognitive Development Lab, and the editors of this volume, Christian Miller and William Fleeson, for their comments on an earlier version of this chapter. I would also like to thank the participants and audience at the Psychology of Character Workshops for their feedback on the research presented in this chapter. This paper was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology 44*, 113–126.

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, *446*, 794–796.

Deutsch, M. (1975). Equity, equality and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, *31*, 137–149.

Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological science*, *24*(6), 860–868.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature 454*, 1079–1083.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–791.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*,137–140.

Freud, S. (1901/1966). *The psychopathology of everyday life* (No. 611). New York: W. W. Norton & Company.

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental science*, 14(5), 1012-1020.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development 26*(1), 30–39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature 450*(7169), 557–559.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences*, *108*(50), 19931–19936.

Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C., . . . Henrich, N. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences 28*, 795–815.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Hobbes, T. (1651/1928). *Leviathan, or the matter, forme and power of a commonwealth ecclesiasticall and civil*. New Haven, CT: Yale University Press.

Hoffman, M. L. (2001). *Empathy and moral development: Implications for caring and justice*. New York, NY: Cambridge University Press.

Homiak, M. (2011). Moral character. In *The Stanford Encyclopedia of Philosophy*. E. N. Zalta, (Ed.).

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science 320*(5879), 1092–1095.

Johnson, T., Dawes, C., Fowler, J., McElreath, R., & Smirnov, O. (2009). The role of egalitarian motives in altruistic punishment. *Economics Letters*, 102, 192–194.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., . . . Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science 8*(6), F31–F36.

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race?. *Cognition*, 119(1), 1-9.

Lind, E. A., & Tyler, T. R. (1988). The social psychology of procedural justice. New York: Springer.

Liszkowski, U. (2005). Human twelve-month-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. *Gesture*, *5*(1–2), 1–2.

Locke, J. (1690/1970). *An essay concerning human understanding*. Menston: Scolar Press.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580.

LoBue, V., Nishida, T., Chiong, C., DeLoache, J. S., & Haidt, J. (2011). When getting something good is bad: Even three-year-olds react to inequality. *Social Development*, *20*(1), 154–170.

Mahajan, N., & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition*, *124*(2), 227–233.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80.

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*(1), 222–231.

Piaget, J. (1953). The origin of intelligence in the child. London: Routledge & Kegan Paul.

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition*, *12*(4), 717–731.

Rousseau, J.-J. (1762/1979). *Emile; or, on Education*. Translated by Allan Bloom. New York: Basic.

Schmidt, M. F., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS One*, *6*(10), e23223.

Shin, S., Burns, M. P.,Yun, J., and Sommerville, J. A. (2011). *When a friend doesn't look like me: 12-month-old infants overcome race bias and prefer social individuals*. Poster session presented at the biennial meeting of the Society for Research and Child Development,. Montreal, Canada.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness?. *Psychological Science*, 23(2), 196-204.

Sommerville, J. A., et al. (in preparation). Correlations between parental empathy and infants' fairness sensitivity.

Sommerville, J. A., Schmidt, M. F., Yun, J. E., & Burns, M. (2013). The development of fairness expectations and prosocial behavior in the second year of life. *Infancy*, *18*(1), 40–66.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*(1), 431–445.

Svetlova, M., Nichols, S. R., & Brownell, C. A. (2010). Toddlers' prosocial behavior: From instrumental to empathic to altruistic helping. *Child Development*, *81*(6), 1814–1827.

Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young children share the spoils after collaboration. *Psychological Science*, *22*(2), 267–273.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*(3), 271–294.

Woodward, A. L. (2013). Infant foundations of intentional understanding. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World: A Developmental Perspective* (pp. 75–80). Oxford: Oxford University Press.

Wright, S. A., Dinsmore, J. B., & Kellaris, J. J. (2013). How group loyalties shape ethical judgment and punishment preferences. *Psychology & Marketing*, *30*(3), 203–210.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*(1), 126.

Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences*, 108, 19761–19766.

Ziv, T. & Sommerville, J. A. (under review). Developmental changes in infants' fairness expectations from 6 to 15 months of age.

21 }

Character Development in the School Years: Relations among Theory of Mind, Moral Identity, and Positive and Negative Behavior toward Peers

Elizabeth A. Boerger and Anthony J Hoffman

The years from age 5 through age 12 are seen as a time when children's development is increasingly influenced by their experiences in school. These years are often referred to as "the school years" to reflect both the amount of time that children spend in school and the importance of school-based experiences for children's cognitive and social-emotional development. The school years are also seen as a crucial period for children's character development, including aspects of character related to achievement (e.g., self-awareness, self-discipline), citizenship (e.g., respect for social standards of behavior), and morality (e.g., respect for the rights and well-being of others) (Damon, 1995; Ji, DuBois, & Flay, 2013). Development of these character strengths depends on age-related changes in children's understanding of themselves and others (e.g., increased ability to perceive themselves as individuals with specific, enduring interests and abilities, which may be different from the interests and abilities of their peers), and on children's immersion in a social context, the school, which provides a wealth of opportunities for discovering and developing their own interests and abilities, for comparing themselves to children with different interests and abilities, and for coordinating their behavior with others' to achieve individual and group goals.

In school, children encounter multiple opportunities for learning about themselves as individuals and as members of groups (Harris, 1995; Harter, 1999). These opportunities occur in both formal instructional contexts that are designed and implemented by adults (e.g., classroom instruction or small group projects), and in informal peer interactions that occur at lunchtime, during recess, and any moments when the teacher's back is turned. Although

both types of group experiences contribute to each of the three broad aspects of character development (i.e., achievement, citizenship, morality), many developmental psychologists consider interactions among peers to be particularly important to moral development during this age period (e.g., Damon, 1988; Keller, 1984; Piaget, 1932/1965; Siegal, 1982; Wainryb, Brehl, & Matwin, 2005). This is because these interactions provide children with their first experiences of independently establishing and maintaining reciprocal relationships (i.e., relationships in which each child is motivated to act in accordance with the other's needs in order to maintain the relationship). To be successful in reciprocal peer relationships, children must be able to organize and participate in joint activities, share resources, and resolve conflicts and to do so in a manner that is seen as fair and rewarding by all parties. However, school-age children are involved in multiple types of peer relationships simultaneously, including best friendships, small group friendships, and relations among a whole class or multiple classes of children from the same grade level. The heterogeneity of peer relationships means that in addition to learning skills associated with maintaining close reciprocal relationships, children also gain experience in creating and defending the boundaries of distinct subgroups among peers, which may involve negative behaviors intended to exclude others (Harris, 1995; Kyratzis, 2004). As a result, peer interactions provide children with frequent and diverse opportunities to engage in helpful (e.g., sharing resources among a group of friends) and hurtful (e.g., teasing another child) behavior toward others, and for observing the consequences of their behavior on their own emotions, other children's welfare, and their individual and interpersonal goals.

A large body of research from the social domain perspective indicates that children begin to understand the moral implications of helping and hurting others during the preschool years, and their ability to evaluate their own and others' behavior in accordance with moral principles increases rapidly through the school years (Jackson & Tisak, 2001; Smetana, 2006; Wainryb et al., 2005). Despite these advances, however, children differ markedly in the extent to which they engage in prosocial and aggressive behavior toward their peers. In recent years, these differences have been the focus of two distinct research literatures: one focused on the behaviors associated with individual differences in two aspects of children's social status in the peer group (i.e., *social acceptance*, which measures the extent to which the child is liked by peers, and *social prominence*, which measures the extent to which peers perceive the child as popular within the peer group), and the other focused on factors that predict children's predominant role in bullying (i.e., active bully, supporter of the bully, victim, or defender of the victim).

Taken together, the findings from these two literatures clearly indicate that children differ in their motivations for prosocial and aggressive behavior, as well as in the amount of each type of behavior they exhibit toward peers. Specifically, some children are more motivated to be seen as popular by the peer

group as a whole than to be personally liked by individual peers. Although these children may engage in relatively high levels of prosocial behavior toward other popular children, they are also involved in high levels of aggression toward unpopular children, a pattern of behavior that typically enhances their status as popular as they move through the school years (Caravita & Cillessen, 2012; Reijntjes et al., 2013; Rodkin & Roisman, 2010; Troop-Gordon, Visonti, & Kuntz, 2011). Children who display this pattern of aggressive behavior are referred to as *popular-aggressive* children (or "popular bullies") to differentiate them from *rejected-aggressive* children, whose aggression is usually associated with both low social acceptance and low social prominence (Garandeau, Lee, & Salmivalli, 2014; Reijntjes et al., 2013).

The popular-aggressive pattern of peer aggression is particularly worrisome because it is resistant to anti-bullying interventions during elementary school (Garandeau et al., 2014), and associated with maintaining a high level of aggression following the transition to junior high school (Cillessen & Mayeux, 2004; Peeters, Cillessen, & Scholte, 2010; Shi & Xie, 2012). This suggests that identifying the factors that underlie individual differences in prosocial and aggressive behavior during the school years is an important step toward the development of effective school-based interventions to divert children from this developmental pathway.

In this chapter, we provide an overview of the research on individual differences in children's peer relationships, including recent work on the social dynamics of bullying. We next review literature on moral development in the school years that highlights children's understanding of the moral implications of their behavior toward peers. Against this background, we review literature on two factors that have been proposed to explain individual differences in children's and adolescents' involvement in prosocial and aggressive behavior: children's theory of mind (i.e., their ability to interpret social interactions in terms of the beliefs and intentions of the individuals involved in the interaction) and moral identity (i.e., the extent to which they have integrated moral knowledge into their sense of self and use it to guide their behavior). Finally, we discuss how moral identity may interact with theory of mind to predispose children to either a predominantly prosocial or predominantly aggressive orientation toward peers, and discuss recent findings from our own lab that support the hypothesized interaction.

Individual Differences in Positive and Negative Behavior toward Peers

For many children, the availability of a large number of same-age children with whom to form friendships and supportive relationships is one of the most attractive features of school. Beginning in kindergarten and continuing

through elementary school, children who are generally liked by their peers and who perceive their peers as supportive report higher levels of liking school than less well-liked children (Boulton, Don, & Boulton, 2011; Ladd, Kochenderfer, & Coleman, 1996). Conversely, children who experience problems in peer relationships, including low acceptance by peers and aggression toward peers, are at greater risk for dropping out of school and for becoming involved in criminal activities later in life (Parker & Asher, 1987). Therefore, understanding the characteristics that differentiate children who are and are not successful in their peer relationships has been of great interest to psychologists studying children's adjustment to school. This research provides considerable insight into children's individual styles of relating to peers, which differ in terms of both the amount of prosocial behavior and aggression children direct toward their peers, and the extent to which aggression is used to attain status within the group.

INSIGHTS FROM STUDIES OF PEER RELATIONSHIPS

Research on individual differences in peer relationships initially focused on peer acceptance, which is usually assessed either by asking children to name "most liked" and "least liked" classmates (Coie, Dodge, & Coppotelli, 1982), or by asking children to individually rate how much they like and dislike each child in the class (Parker & Asher, 1993). In either case, children's status is determined by calculating a score that indicates both how liked and disliked each child is relative to other members of the class. Although these measures typically reveal several distinct patterns of peer acceptance, the most marked differences in both behavior and developmental outcomes are between highly accepted children, who are liked by many and disliked by few classmates, and rejected children, who are liked by few classmates and disliked by many (Coie et al., 1982; Parker & Asher, 1987, 1993). In general, highly accepted children are seen as physically attractive and socially skilled. They engage in high levels of prosocial behavior toward peers, display little aggression, and are seen as leaders among their peers. Rejected children show the opposite pattern of characteristics. They are rated as aggressive, disruptive, and high in help-seeking behavior. Furthermore, peer rejection, especially of children who are seen as both aggressive and withdrawn, is a relatively stable aspect of children's school experience (Ladd & Burgess, 1999).

The close connection between aggression and peer rejection in studies that measured peer status solely in terms of peer acceptance led to a widely held belief that peer aggression primarily involved socially marginalized children with poor social skills (Crick & Dodge, 1994). However, both longitudinal and cross-sectional studies suggest that the rejected group encompasses children with diverse characteristics, which may be associated with different developmental trajectories (Farmer, Hall, Leung, Estell, & Brooks, 2011; Parker &

Asher, 1987). In order to better understand diversity in the quality of peer relationship, psychologists have begun to study peer acceptance in conjunction with the second dimension of peer status: the child's level of social prominence or popularity. This is measured by asking children to name the popular and unpopular students in their class. Although there is some overlap between children who are rated as popular and children who are highly accepted by peers, popularity and peer acceptance are separable aspects of children's social status. Thus, popular children can be divided into two groups who differ in terms of their behavior as well as their level of peer acceptance: popular-prosocial children who are predominantly known for their prosocial behavior (these children are usually also high in peer acceptance), and popular-aggressive children (these children are usually average to moderately high in peer acceptance, but may be rejected). Regardless of their level of peer acceptance, popular-aggressive children are often seen not only as bullies but also as leaders in the classroom. The fact that these children are often admired even by students who dislike them personally can contribute to a class environment in which children believe that bullying is accepted by the majority of their peers (Farmer et al., 2011; Reijntjes et al., 2013; Troop-Gordon et al., 2011).

INSIGHTS FROM STUDIES OF BULLYING

In recent years, a burgeoning literature on the social dynamics of bullying has revealed much about the role of children's perceptions of group norms with regard to aggression in the maintenance of either a positive or a negative peer culture. Although relatively few children are directly involved in bullying as either bullies or victims, a majority of children in classes where bullying is prevalent support the bullies by either actively assisting or reinforcing them or passively accepting their behavior (e.g., Hawkins, Peplar, & Craig, 2001; Sutton, Smith, & Swettenham, 1999). Differences in the frequency and severity of bullying are associated with both the extent to which popular children engage in bullying, and children's beliefs about whether bullying is accepted by the majority of their peers (Salmivalli, 2010). Although most children personally disapprove of bullying and dislike bullies, children tend to believe that the peer group as a whole thinks that bullying is acceptable. A particularly disturbing finding is that many children who underestimate the extent to which their peers share their disapproval of bullying respond to bullying by joining in, rather than by attempting to stop the incident or assist the victim (Sandstrom, Makover, & Bartini, 2013). Children's responses to bullying are also highly influenced by the bully's choice of victim. They are most likely to express attitudes consistent with moral disengagement, for example, that the victims deserve their mistreatment, when bullying is performed by popular classmates against one or two socially isolated peers rather than when aggression is directed against many different children. Furthermore, repeated victimization of the same

children increases the extent to which victims view themselves as lacking in self-worth and reduces the victims' level of peer acceptance even among their defenders (Salmivalli, 2010).

IMPLICATIONS FOR MORAL DEVELOPMENT

The findings from the peer relationships and bullying literature highlight the complexity of the social and moral decisions children face in their day-to-day peer interactions. In addition to making straightforward moral decisions about their own behavior toward one or two other children (e.g., whether to push Tommy off the swing or wait patiently for a turn), children have to respond to the behavior of all the other children in their class or on the playground. Therefore, their decisions will inevitably be influenced by their observation of other children's behavior, as well as by their beliefs about how their own behavior will affect an assortment of personal and social goals, including goals related to gaining friends or achieving prominence within the group, or even for protecting their own safety.

This analysis suggests that children need to coordinate three types of knowledge in order to develop and maintain a predominantly prosocial style of interacting with peers. These include (1) moral knowledge, or their understanding of moral principles that should guide their decisions about the right thing to do; (2) social-cognitive knowledge, or their ability to interpret social events and to choose effective behaviors to achieve desired outcomes, which may be defined in terms of personal, interpersonal, or moral concerns; and (3) their sense of moral agency, or the extent to which they understand relations between their intentions and behavior, feel responsible for the effects of their behavior, and feel obligated to behave in accordance with moral rules. Of these, moral knowledge is no doubt the most important. Children cannot behave morally unless they are aware of moral principles and understand that they define obligatory behavior. However, social-cognitive ability is also important because it enables children to have social relationships with peers. After all, it is only when children are engaged with their peers that they can behave morally toward them. Finally, a sense of moral agency is required to enable children to coordinate their moral knowledge and their social-cognitive knowledge in order to regulate and evaluate their behavior in their interactions with peers.

In recent years, development in children's social-cognitive knowledge has primarily been studied in terms of a construct called *theory of mind* (ToM). This term, which is used similarly to the term "folk psychology" in philosophy, refers to children's ability to integrate information about relations between subjective states (i.e., desires, beliefs, and intentions) and behavior, into an implicit "folk theory" that guides their social responses. The theory enables children to interpret their own behavior, as well as the behavior of other people, in terms

of the desires, beliefs and intentions that guided it, and to use this information to coordinate their behavior with others in social interactions. Therefore, our references to ToM refer to children's social-cognitive knowledge.

Although there has been little research on development in moral agency during the school years, development in this aspect of morality is studied in terms of the development of a moral identity in adolescence. According to Hardy and Carlo (2011), an adolescent's ability to form a moral identity depends on an earlier development in the child's understanding of the self as an individual who is responsible for the good and bad effects of his or her behavior across time. Moral identity incorporates this earlier sense of moral agency but adds the component of a self-chosen commitment to behave in a manner consistent with moral principles rather than personal or social goals, when those goals conflict with moral principles. Therefore, we use *moral agency* to refer to children's initial ability to independently evaluate their behavior in terms of moral principles and hold themselves responsible for their good and bad acts, and *moral identity* to refer to an elaboration of this initial sense of moral agency that includes a commitment to prioritizing moral principles when these are in conflict with other goals.

Moral Development in the Preschool and School Years

Research over the past 40 years has established that even very young children are keenly aware of moral and immoral actions that occur in the context of their everyday social interactions, and can differentiate between moral violations (e.g., hitting another child) and other types of misbehavior (e.g., not picking up toys). Much of this research has been conducted from the perspective of social domain theory and relates to children's reasoning about behavior that is permitted or forbidden in various social domains, including the moral, conventional, and personal domains (Smetana, 2006). According to social domain theory, children begin to develop concepts of distinct social domains, which are regulated by different types of rules, through social interactions surrounding rule violations in toddlerhood and early childhood (Nucci & Weber, 1995; Smetana, Kochanska, & Chuang, 2000; Smetana et al., 2012). By 3 to 4 years of age, children differentiate between moral rules, which, for example, prohibit deliberately hurting others, and conventional rules, which prescribe behaviors that promote orderly social interaction, and understand that moral rules are "stronger" than conventional rules. Specifically, young children understand that moral rules apply to parents and teachers as well as to children, that they apply at home as well as at school, and that the actions they prohibit would be wrong even if there were no rules about them in place. In contrast, children know that conventional rules may be specific to the behavior of some individuals (e.g., children but not adults), can vary across contexts,

and that the behaviors regulated by these rules would not necessarily be wrong if there were no rules about them in place. During the preschool years, children also become aware that some behaviors fall into the personal domain, meaning that they are not usually regulated by rules (Nucci & Smetana, 1996; Nucci & Weber, 1995). Therefore, preschoolers, like older children and adolescents, frequently argue with parents who attempt to assert authority over decisions—for example, about whom to play with or what to play—that children see as belonging in the personal domain.

As children make the transition from preschool to school age, their understanding of the moral domain expands, and they become increasingly able to reason about situations that involve overlap between domains (Smetana, 2006). In particular, as children gain experience with the complex forms of peer interactions that occur in school, they become increasingly aware of the harm caused by psychological and relational forms of aggression, such as teasing, social exclusion, and coercion (e.g., unfair demands for "sharing" personal property, as when one child demands another child's cookie), and expand their concept of the moral domain to include these behaviors (Murray-Close, Crick, & Gallotti, 2006; Park & Killen, 2010; Shaw & Wainryb, 2006). Children also begin to reason about factors that either increase (e.g., lying in addition to stealing) or reduce (e.g., failing to share candy with a disliked peer vs. failing to give him lunch) the severity of moral transgressions (Smetana, Killen, & Turiel, 1991), and consider contextual factors that affect whether a particular behavior represents a moral transgression or should be considered a matter of personal choice (Park & Killen, 2010; Wainryb et al., 2005). For example, Park and Killen found that American and South Korean fifth graders, like their counterparts in eighth grade, considered it more acceptable to reject a child as a friend than to exclude him or her from a club, which, in turn, was seen as more acceptable than actively excluding the child from peer interactions by regularly teasing and picking on the child. Although the last form of exclusion was considered unacceptable regardless of the victim's characteristics, children's judgments about the less severe forms varied depending on the reasons given for rejecting or excluding the child. Specifically, children considered it more acceptable to reject a child as a friend or club member on the basis of personal characteristics (e.g., aggression) than on the basis of group membership (e.g., nationality).

School children also begin to integrate their understanding of moral rules with their understanding of emotions, and to reason about the emotional consequences of moral transgressions for the transgressor as well as the victim. Specifically, whereas preschoolers typically predict that they would feel happy after breaking a moral rule if it allowed them to achieve a desired end (a response referred to as the *happy victimizer* phenomenon), children increasingly come to anticipate that they would feel bad or "a little bit bad and a little bit good" by 6 or 7 years of age (Krettenauer, Malti, & Sokol, 2008;

Nunner-Winkler & Sodian, 1988). This shift from a "happy victimizer" perspective to "guilty victimizer" perspective is seen as an indication that children not only understand moral rules, but recognize them as personally binding. Thus, children's attributions of negative emotions to the self-as-transgressor in response to hypothetical stories are believed to reflect the extent to which they experience these emotions following actual transgressions. Children's ability to anticipate not only the valence but also the intensity of the moral emotions they would feel after fulfilling or disregarding a moral obligation appears to increase from late childhood through adolescence, and may help to motivate children toward moral behavior (Krettenauer & Eichler, 2006).

Children's ability to reflect on and evaluate their own harmful behavior toward peers also increases rapidly during the school years. These changes are reflected in children's narrative reports of situations in which they "did or said something and a friend or another child that you know well ended up feeling hurt by it" (Wainryb et al., 2005, p. 23). Although children of all ages include references to the negative emotion experienced by the child they harmed in their narratives, older children's narratives show much greater understanding of the psychological aspects of the interaction that preceded their harmful act. In contrast to preschoolers' narratives, which focus narrowly on the narrator's actions and the victim's emotion (e.g., "I said, 'I don't want to play with you,' and he was sad"), the narratives of first graders include references to the conflicting desires of the two children (e.g., "My friend wanted to paint and I really didn't want to paint"), and fifth graders' narratives include discussion of differences in what each of the two children thought during the conflict as well as differences in what they wanted (e.g., "I thought she expected me to be the problem-solver, but I'm not. I hate it. But she couldn't know that I felt that way, right?").[1] Thus, in contrast to the preschoolers' narratives, which suggest they interpret these events as sequences of behavior that are unrelated to their own, as well as their victim's, intentions, the older children's narratives show evidence of an attempt to coordinate their own and the other child's intentions as a way to understand their transgression.[2] The greater psychological complexity of

[1] These examples are based on children's narratives included in Wainryb et al., 2005, pp. 55–58.

[2] Wainryb et al. (2005) asked children to discuss two events that happened with a friend or other well-known peer: one in which the narrator was hurt by the other child, and one in which he or she hurt the other child. Therefore, each child provided one "victim narrative" and one "perpetrator narrative." The results discussed here refer specifically to age-related changes in children's perpetrator narratives. Similar age-related changes in references to mental states occurred in the victim narratives, but across age groups (up to age 16), victim narratives focused on the child's own perspective, whereas perpetrator narratives presented the event from the perspective of both children. As a result, the perpetrator narratives, especially those of the school-age and high school participants, were less coherent than the victim narratives, with the point of view shifting frequently from the narrator's perspective to his or her inferences about the other child's perspective.

the older children's narratives is reflected in their moral evaluations of their behavior. Whereas preschoolers typically evaluate their behavior as categorically wrong (or, less frequently, categorically right), school-age children are increasingly able to balance concern with the harm experienced by the victim with their understanding of their own intentions and to judge their behavior more harshly when it is motivated by hostile or selfish intentions (Recchia, Wainryb, & Pasupathi, 2013; Wainryb et al., 2005).

Taken together, the findings reviewed in this section indicate that school-age children have at least a nascent understanding of themselves as moral agents. That is, as children move through the school years, they are increasingly able to interpret interactions with friends and peers as situations involving overlap between the personal and moral domains and to evaluate their own hurtful behavior toward peers in terms of both personal rights and moral obligations. They are also increasingly able to anticipate that seeking personal gain by harming a peer will have negative emotional consequences for the transgressor as well as the victim. Despite these developing abilities, however, children differ in the extent to which their behavior toward peers reflects concern with moral principles. As noted previously, these differences may reflect differences in children's theory of mind ability (i.e., their ability to interpret and respond appropriately to other children's behavior), their sense of moral agency or moral identity (i.e., the extent to which they are intrinsically motivated to behave morally), or an interaction between these two aspects of self-knowledge and social knowledge. Therefore, we next turn to an examination of research that has explored relations between these factors and peer relationships among school-age children.[3]

Relations among Theory of Mind, Moral Identity, and Positive and Negative Peer Behavior

ToM refers to children's understanding of the relations between real-world events and the subjective desires, thoughts, and emotions that guide people's actions with respect to these events. This knowledge is referred to as a "theory" because it is believed that children, as well as adults, organize their understanding of mental states into an informal explanatory framework

[3] To the best of our knowledge, no previous studies have studied moral identity in school-age children. This does not mean, however, that there are no studies that assess the moral motivation as a factor in children's peer interactions. Rather, in the studies we review, moral motivation is assessed through measures of children's moral emotion expectancies (i.e., their attribution of negative emotions to moral transgressors) and measures of moral disengagement (i.e., the extent to which they endorse statements that minimize a transgressor's responsibility for an immoral act).

that allows them to interpret and respond to complex social situations (Flavell, 2004). Although children's understanding of others' mental states has roots in children's experiences with their parents and siblings in infancy and toddlerhood (Dunn, 1988), it appears to undergo a fundamental shift between 4 and 6 years of age, when children acquire an understanding of false beliefs (Wellman, Cross, & Watson, 2001). This ability allows children to infer when others will have false beliefs about objective events and to predict how individuals with false beliefs will act in relation to the objects represented by those beliefs. For example, in one famous task used to test children's false belief understanding, a child is shown a scene involving two puppets. The first puppet (Maxi) has a chocolate bar that he puts into one cabinet before going outside to play. Then the second puppet (Maxi's mother) removes the chocolate bar from the first cabinet to a second cabinet, and leaves the room. Finally, Maxi returns, and the child is told that Maxi is now hungry and wants to eat his chocolate bar. The child is asked to report where the chocolate bar really is and where Maxi will look for the chocolate bar. Children who correctly answer that the chocolate bar is in the second cabinet and that Maxi will look for it in the first cabinet are credited with understanding that Maxi falsely believes that his chocolate bar is still in the cabinet where he left it (Wimmer & Perner, 1983).This new understanding of the relations between real world events, beliefs, and behavior allows children to begin to interpret their own experiences as events that are coherently structured in terms of the intentions of the individuals involved, and increases children's ability to coordinate their behavior with others in social interactions involving one or two other children. However, being able to effectively manage social relationships with a large number of other children also requires an ability to predict how one's own behavior will affect others in unfamiliar and ambiguous situations. Making accurate predictions about the effects of one's own behavior on other children in these situations also requires knowledge about the effect of belief on emotions and the role of recursive perspective taking (i.e., one individual's belief about another's belief) in social interactions. These more advanced ToM abilities develop rapidly during the school years, and culminate in children's ability to predict and explain others' behavior and emotional responses in terms of so-called higher-order beliefs (Liddle & Nettle, 2006). These include not only second-order beliefs (e.g., "John thinks that Mary wants him to hold her hand") but also third- and fourth-order beliefs (e.g., "John thinks that Mary wants him to believe that she wants him to hold her hand."). Mastery of higher-order beliefs allows children to comprehend speech acts, such as promises, that depend on the sharing of intentions about future events (Mant & Perner, 1988) and is associated with teachers' ratings of children's adjustment to the academic and social demands of school among 10- and 11-year-olds (Liddle & Nettle, 2006).

CONFLICTING ACCOUNTS OF THE EFFECT OF HIGHLY DEVELOPED TOM ON BEHAVIOR TOWARD PEERS

Although much research on ToM in typically developing children has focused on charting the sequence of the age-related developments described in the previous paragraph, it is also apparent that there are individual differences in when children master these abilities (Hughes & Leekam, 2004). This has led to debate about the effect of these differences on children's ability to effectively manage social relationships, particularly peer-group interactions. Two theoretical accounts have guided research on the relation between ToM and children's adjustment to the peer social context. These accounts differ in the extent to which they predict that increases in ToM abilities will lead to more positive behavior toward peers.

According to the first account, which is known as the *social-constructivist view*, children's elaboration of their ToM during middle childhood depends on their involvement in diverse social relationships. Through interactions with adults, older children, and peers, school-age children have opportunities to encounter diverse interpretations of the same events and to gain experience in articulating their own point of view and modifying it in response to views expressed by others. A central expectation with respect to the social-constructivist view is that children who have more ability to recognize others' perspectives as different from their own will be more attractive as interaction partners, and will therefore have more opportunities for discussion and debate, which will, in turn, promote further development of their ToM (Banerjee, Watling, & Caputi, 2011; Peterson & Siegal, 2002). Therefore, children who begin school with more advanced ToM abilities are expected to enjoy increasing peer acceptance over the school years; whereas children who have fewer ToM abilities are expected to be at increasing risk for peer rejection.

Two recent longitudinal studies conducted by Banerjee and Caputi and their colleagues provide support for this perspective. In the first of these studies, Banerjee et al. (2011) studied the bidirectional effects of ToM and continuous measures of peer acceptance (i.e., number of "most liked" nominations) and peer rejection (i.e., number of "most disliked" nominations) over a two-year period. Results showed that peer rejection early in the study predicted poor performance on the ToM task relative to peers one year later, which, in turn, predicted higher peer rejection the following year. In contrast, among the older children, higher ToM scores at the beginning of the study were associated with greater peer acceptance one year later. In a separate study, Caputi, Lecce, Pagnin, and Banerjee (2012) followed a group of children from their last year of preschool through their second year of elementary school (equivalent to first grade in the United States). Children who had more advanced ToM abilities during the last year of preschool engaged in more prosocial behavior to peers in elementary school. Furthermore, both the children's tendency toward

prosocial behavior and their understanding of others' thoughts and emotions contributed to being liked by peers two years later. These studies provide the best evidence to date that early-developing ToM abilities promote adjustment to the peer group, which in turn facilitates the development of more complex understandings of relations between others' thoughts and emotions and their behavior. Nonetheless, because measures of negative social behaviors (e.g., aggression toward peers) were not included in either study, the results do not clearly indicate that children who have advanced ToM understanding invariably employ it toward prosocial ends.

This is a serious limitation because, although ToM abilities may make children more sensitive to others' needs, they can also increase children's awareness of their own power to manipulate others' beliefs and behaviors to promote their own social goals. A propensity to use ToM abilities toward negative ends may be especially likely among children who experience little empathy toward others (Sutton, 2003; Sutton et al. 1999). This idea is central to the second theoretical account of relations between children's developing theories of mind and their peer relationships, which might be referred to as a *social goals perspective*. In putting forth this perspective, Sutton argued that some children who are especially good at predicting how their actions affect their peers' beliefs and behaviors might use this ability to subtly undermine others in the peer group in order to enhance their own status. In particular, Sutton hypothesized that children who were high in ToM but low in empathy might be especially likely to become bullies. In support of this hypothesis, Sutton et al. found that so-called *ringleader bullies*, or children who appear to promote bullying among other children, performed better on ToM tasks than any other group of children, including children who defended the victims. Thus, in this study, ToM appeared to be more closely associated with negative peer behavior than with positive peer behavior. However, Sutton et al.'s study did not provide any evidence about the hypothesized link between low empathy and bullying among children with well-developed ToM abilities. Their study thus provides little insight into why children might choose to use their ability to understand social situations to harm rather than help others.

Caravita, Di Blasio, and Salmivalli (2009) set out to directly test Sutton's (2003) hypothesis that empathy moderates the relation between ToM ability and involvement in bullying. The specific hypothesis they tested was that children who were high in both ToM and empathy would be more likely than other children to defend victims; whereas children who were high in ToM but low in empathy would be more likely to bully others. The results showed that ToM was significantly positively related to both defending behavior (among both boys and girls), and ringleader bullying (among boys). Thus, having highly developed ToM skills appeared to predispose children to active roles in bullying situations, as either bullies or defenders. However, the association between ToM and bullying did not depend on a lack of empathy. Rather, boys who

were high in ToM and either empathy or peer acceptance were more likely to intervene to defend bullied peers.

Taken together these findings clearly show that having the ability to accurately interpret complex social situations in terms of others' desires, beliefs, and intentions does not invariably lead children to engage in positive behavior toward their peers. Although children who have more advanced ToM abilities may have an advantage in terms of making and maintaining friends, as is suggested by Banerjee et al.'s (2011) study, the overall impression left by this set of studies is that having good ToM abilities in comparison to peers is associated with playing a more active role within the peer group, although this role may or may not be a prosocial one.

MORAL MOTIVATION AS A PREDICTOR OF POSITIVE AND NEGATIVE BEHAVIOR TOWARD PEERS

In response to conflicting findings concerning the role of ToM in peer relationships, researchers have begun to consider other factors that may predict prosocial behavior and aggression in children's peer interactions. The results of two recent studies suggest that individual differences in the emotions children expect to experience following a moral transgression may play a role in determining whether children with advanced ToM abilities are more likely to be involved in prosocial behavior or aggression. Specifically, these studies assessed children's responses to several happy victimizer scenarios, in which one child gained something he or she wanted by harming another child. Children's responses were scored as showing high *moral cognition* (i.e., understanding of behaviors that involve moral transgressions) if the child stated that the action was wrong; and, as showing high *moral motivation* if he or she attributed a negative emotion to the transgressor. Among young elementary-school children (7- to 8-year-olds), Gasser and Keller (2009) found that bullies and prosocial children performed similarly well on ToM tasks, but that bullies performed poorly on both moral cognition and moral motivation. Thus, despite their high ToM scores, bullies appeared to be delayed in both their understanding of moral principles and their development of a sense of moral agency. That is, in contrast to most school-age children, the bullies did not interpret hurtful behavior between peers as a violation of a moral rule, nor did they anticipate that the behavior would cause the transgressor to feel guilty, suggesting that they had not internalized either moral principles or an understanding of moral rules as personally binding.

Using a similar method with an older sample (8- to 11-year-olds), Gini (2006) found a somewhat different pattern of results. Specifically, bullies and defenders performed equally well on both ToM and moral motivation questions. The two groups differed, however, in their responses to a moral disengagement questionnaire that assessed their level of agreement with such statements as

"some kids deserve to be treated like animals," which reflect excuses for treating others badly. Specifically, bullies scored higher than all other children on this measure. The high moral motivation scores of these bullies suggest that, in contrast to the younger children studied by Gasser and Keller (2009), they have an internalized understanding of moral obligation that allows them to recognize situations calling for moral action and determine the correct way to behave in such situations. However, their high moral disengagement scores suggest that they have also internalized a cognitive strategy for minimizing their sense of responsibility in circumstances in which they would prefer to pursue other goals. In other words, these bullies may have adopted a moral identity in which values associated with pursuing personal gain or social prominence are more central than moral principles. Thus the differences in behavior between the high-ToM children who became bullies and the similarly high-ToM defenders may reflect differences in the moral identities adopted by the two groups of children. In support of this interpretation, the bullies' use of moral disengagement as a strategy to avoid feeling morally responsible for others' well-being is consistent with findings that adolescents and adults who score low on measures of moral internalization (one component of a moral identity) are more likely to engage in moral disengagement in response to stories about human rights abuses, such as the abuse of Iraqi prisoners by US soldiers at Abu Ghraib prison (Aquino, Reed, Thau, & Freeman, 2007).

DO THEORY OF MIND AND MORAL IDENTITY INTERACT TO PRODUCE MORAL BEHAVIOR?

In a study funded by the Character Project, we (Boerger, Mauthe, Falkowitz, & Burns, 2013) examined whether individual differences in ToM and moral identity might together predict differences in school-age children's levels of prosocial behavior and aggression toward peers. According to the model of moral identity that informed the design of the study, individuals adopt a moral identity by integrating a coherent set of beliefs about the traits that define a moral person (e.g., kindness, honesty, responsibility) into their self-concept and use these beliefs to guide their behavior in morally relevant situations. (Aquino & Reed, 2002; Hardy, Bhattacharjee, Reed, & Aquino, 2010). The integration of the beliefs into the self-concept causes the individual to feel that behaving morally is behaving in accordance with his or her own desires. Therefore, individuals who have formed a moral identity will be intrinsically motivated to behave in ways that reflect the set of traits that define a moral person. However, individuals can experience and express their moral identity in two different ways. *Moral internalization* refers to the extent to which behaving morally (i.e., being kind, honest, responsible, etc.) is important to the individual's sense of self. Therefore, individuals who are high in moral internalization are motivated to display behaviors consistent with their view of a moral person even when this

requires costs, such as requiring the individual to give up pursuit of other personally meaningful goals. A second component of moral symbolization refers to the extent to which publically communicating one's commitment to moral ideals is important to one's sense of self. These two aspects of moral identity are associated with distinct patterns of morally relevant attitudes and behavior. Specifically, moral internalization is negatively associated with normlessness (i.e., willingness to engage in behavior that violates social norms in order to meet personal goals) and positively associated with moral reasoning, whereas symbolization is associated with religiosity (Aquino & Reed, 2012).

In our study we were especially interested in exploring relations between individual differences in moral internalization and children's involvement in positive and negative peer relationships. In particular, we hypothesized that the apparently contradictory findings about the role of ToM in children's peer relationships might be explained by differences in the moral internalization component of moral identity. We expected that among children who were high in ToM, those who were also high on moral internalization would be more prosocial and less aggressive in peer interactions than children who scored low on moral internalization.

To test this hypothesis, we individually interviewed third- through seventh-grade children from two Mississippi schools to assess their higher-order ToM ability and moral identity, as well as their perception of their abilities in several social (i.e., ability to make and maintain friends and to regulate behavior) and individual achievement (i.e., academic and athletic ability and physical appearance) domains. In addition, the children's teachers completed surveys about each child's prosocial behavior and aggression toward peers. Moral identity was measured with an adapted version of Aquino and Reed's (2002) moral-identity questionnaire (MIQ), which we read aloud to the children.[4] This questionnaire assessed children's beliefs about their own moral traits (i.e., the extent to which they saw themselves as compassionate,

[4] The MIQ (Aquino & Reed, 2002) is a self-report measure in which participants are given a list of nine moral traits (caring, compassionate, fair, friendly, generous, hardworking, helpful, honest, and kind) and instructed to "visualize in your mind the kind of person who has these traits" and "imagine how that person would, think, feel, and act" (p. 1427). Participants then use a five-point Likert-type scale to rate how much they agree or disagree with five statements assessing how they would feel if they were the imagined person (internalization) and six statements describing behaviors that express the individual's identity as a person with the moral traits (symbolization). To facilitate children's understanding of the scale, we read the trait words and initial instructions, as well as each item, aloud to the child. We also expanded the instructions for visualizing a moral person, and adapted the phrasing of the items so that each original statement was paired with a statement indicating the opposite feeling or behavior. Children were told that they should choose which of the two statements was most like them, and were then asked to indicate whether the selected statement was "a little" or "a lot" like them. The adapted instructions were: "I want you to make a picture in your mind of a kid who can be described with all these words: [list of traits]. Imagine how that kid would think about things, how the kid would feel, and how the kid would act. You can also think about what kind of clothes the kid would wear, what groups the kid would be part of, and what things the kid would like to do outside of school." An

welcoming, caring, honest, fair, and responsible) as well as their level of moral internalization (i.e., their desire to be like a child with many moral traits) and moral symbolization (i.e., the extent to which children considered it important to display their moral traits to others).

Initial results from this study showed significant correlations between children's moral internalization scores and their teachers' ratings of their involvement in prosocial behavior and aggression. Specifically, teachers rated children with high moral internalization scores as more prosocial and less aggressive than other children. However, when we examined these correlations separately for children who scored high and low on the ToM task, we found that the correlations between children's moral internalization scores and teachers' ratings of prosocial behavior and aggression were only significant in the group of children who scored high on the ToM task. Thus, among children with advanced ToM abilities, but not among those with less advanced abilities, children for whom being moral was more important to their sense of self were more prosocial and less aggressive than children for whom morality was less important.[5]

Although these results are preliminary and correlational, they suggest that once children have achieved a relatively advanced level of ToM ability, their beliefs about the kind of person they want to be (i.e., a person defined by moral traits in the case of children with high moral internalization) begin to motivate behavior that is consistent with this self-view. These results are intriguing, but additional research is necessary to clarify exactly how ToM and moral internalization contribute to individual differences in prosocial and aggressive behavior. One possibility is that being able to interpret social interactions in terms of what each individual knows about the cognitive and affective states of the others

---

example of an adapted item is the question, "Which of these statements is like you, 'It would make me feel good to be like this kid' or 'It would make me feel bad to be like this kid.' Is that a little like you, or a lot like you?" This format has been used frequently in studies of children's self-knowledge (e.g., Harter, 1985; Jackson & Tisak, 2001).

To our knowledge, this measure has not previously been used with preadolescent children. However, the results of our study provide initial support for the validity of the measure, at least with respect to internalization items. Specifically, a principle components factor analysis returned a two-factor solution, with each of the items loading onto the same factor (internalization or symbolization) as in Aquino and Reed's initial study (2002, pp. 1427–1428). In addition, children's internalization factor scores were significantly correlated in the expected direction with self-rated behavioral conduct and moral traits, but were not correlated with self-rated physical appearance social acceptance, academic competence, or athletic competence. Symbolization scores were not significantly correlated with any measure of children's self-concept, which is consistent with the idea that symbolization refers to the extent to which individuals are motivated to display their commitment to moral values in their public behavior. However, these scores were also not related to a measure of altruism (i.e., donation to a charity), which is seen as a form of symbolic expression of one's moral values.

[5] It is important to note that the groups of children defined in terms of ToM ability did not differ with respect to age (both groups had an average age of 10 years and included children from each grade level). Therefore, the differences in the pattern of correlations between moral internalization and behavior in the high and low ToM groups cannot be explained as an effect of age.

increases children's ability to differentiate between moral and expedient actions in their peer-group interactions. For example, children with more advanced ToM ability may be less likely to erroneously assume that their peers approve of bullying when they observe other children reinforcing a bully. Rather, they may correctly infer that these children are afraid of the bully or uncertain about how to respond, and may recognize that creating such feelings of fear and uncertainty is one of the bully's aims. Such insights may help children to frame the incident as one that involves a choice between acting morally (i.e., standing up for the victim) or expediently (i.e., going along with the crowd) and to choose the behavior that is consistent with their internalized moral identity. If so, children who are high in moral internalization, meaning that having moral traits is highly important to their sense of self, would be motivated to stand up for the victim in order to maintain consistency with their self-view. In contrast, children who are low in moral internalization, meaning that having moral traits is less important to their self-view, would be less motivated to stand up for the victim. For these children, other motivations, such as a desire to win favor with the bully, would be more likely to influence their behavior.

In the scenario just described, moral internalization promotes moral behavior because of the child's desire to maintain a consistent sense of self. In other words, children with higher moral internalization scores would be more likely to help the bullied child because failing to do so would conflict with their self-image as a compassionate person or as a person committed to justice. However, the negative emotions associated with failing to live up to one's self-view are not the only negative emotions that a child for whom being moral is important is likely to feel if she fails to help a bullied child. This child will see herself as morally obligated to help the child who is being harmed, and if she fails to do so, she will expect to experience intense guilt. In contrast, another child, for whom being moral is less important, may perceive the moral obligation inherent in the situation but minimize her emotional response to it by deciding that the bullied child is responsible for the attack. Thus, another possibility is that children with more advanced ToM abilities are better at anticipating how their own decisions (e.g., to intervene to stop the bullying vs. joining in) will affect their self-evaluations and moral emotions, therefore increasing their desire to act in accordance with their moral beliefs (if they are high in moral internalization), or to morally disengage (if they are low in moral internalization).

Summary and Conclusion

In this chapter we have argued that children's peer interactions, especially those that take place in school, are an important context for the development of moral character. Within this context children have their first experience of the challenges involved in independently initiating and maintaining

reciprocal relationships; of shifting between close friendships and group relationships, and of regulating behavior appropriately within each type of relationship. These relationships create numerous opportunities for children to engage in both prosocial behaviors and aggression toward peers. Children's success in the peer group, as measured by their overall level of peer acceptance, depends largely on whether their behavior is viewed by peers as predominantly prosocial rather than aggressive. Peer acceptance, however, is not the only important dimension of status within the peer group. Children also differ in terms of their social prominence (i.e., popularity) and some children appear to value this aspect of peer status more than they value peer acceptance. In contrast to peer acceptance, social prominence depends less on prosocial behavior and more on the strategic use of aggression as a means of demonstrating dominance.

Research reviewed in the chapter indicates that early developments in children's understanding of others' thoughts and emotions may support the development of prosocial behavior in school. Most notably, Caputi et al. (2012) have shown that better performance on ToM tasks during preschool predicts more prosocial behavior during children's first school year, and more peer acceptance during their second year. Other studies, however, show that ToM can also promote negative behavior that may help children achieve or maintain social prominence. Therefore, understanding how children's developing ToM abilities interact with other aspects of their social and moral development is crucial to developing effective school-based character development and anti-bullying programs. Results reported in this chapter suggest that advances in ToM ability in late childhood, particularly increases in children's ability to coordinate inferences about the mental states of multiple individuals involved in a social interaction, may facilitate children's ability to regulate their behavior in accordance with internalized moral values. Future research efforts should be directed toward clarifying the exact relationship between ToM and the development of an internalized moral identity.[6]

References

Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440. doi:10.1037/0022-3514.83.6.1423

[6] This publication was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation to Elizabeth Boerger. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Character Project, Wake Forest University or the John Templeton Foundation. Special thanks go Mike Furr and Christian Miller for helpful comments on a previous draft.

Aquino, K., Reed, A. II, Thau, S., & Freeman, D. (2007). A grotesque and dark beauty: How moral identity and mechanisms of moral disengagement influence cognitive and emotional reactions to war. *Journal of Experimental Social Psychology*, *43*, 385–392. doi:10.1016/j.jesp.2006.05.013

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidirectional associations. *Child Development*, *82*, 1887–1905. doi:10.1111/j.1467-8624.2011.01669.x

Boerger, E. A., Mauthe, K., Falkowitz, S., & Burns, A. (2013, October). *Moral identity is related to self-perceived and teacher-rated behavioral conduct in 3rd-7th grade children*. Poster presented at the biennial meeting of the Cognitive Development Society. Memphis, TN.

Boulton, M. J., Don, J., & Boulton, L. (2011). Predicting children's liking of school from their peer relationships. *Social Psychology of Education*, *14*, 489–501. doi:10.1007/s11218-011-9156-0.

Caputi, M., Lecce, S., Pagnin, A., & Banerjee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology*, *48*, 257–270. doi:10.1037/a0025402

Caravita, S. C. S., & Cillessen, A. H. N. (2012). Agentic or communal? Associations between interpersonal goals, popularity, and bullying in middle childhood and early adolescence. *Social Development*, *21*, 376–395. doi:10.1111/j.1467-9507.2011.00632.x

Caravita, S. C. S., Di Blasio, P., & Salmivalli, C. (2009). Early adolescents' participation in bullying: Is ToM involved? *The Journal of Early Adolescence*, *30*, 138–170. doi:10.1177/0272431609342983.

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development*, *75*, 147–163. doi:10.1111/j.1467-8624.2004.00660.x.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*, 557–570.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social-information processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101.

Damon, W. (1988). *The moral child: Nurturing children's natural moral growth.* New York, NY: Free Press.

Damon, W. (1995) *Greater expectations: Overcoming the culture of indulgence in our homes and schools.* New York, NY: Free Press Paperbacks.

Dunn, J. (1988). *The beginnings of social understanding.* Cambridge, MA: Harvard University Press.

Farmer, T. W., Hall, C. M., Leung, M., Estell, D. B., & Brooks, D. (2011). Social prominence and the heterogeneity of rejected status in late elementary school. *School Psychology Quarterly*, *26*, 260–274. doi:10.1037//a00025624.

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, *50*, 274–290. doi:10.1353/mpq.2004.0018.

Garandeau, C. F., Lee, I. A., & Salmivalli, C. (2014). Differential effects of the KiVa anti-bullying program on popular and unpopular bullies. *Journal of Applied Developmental Psychology*, *35*, 44–50. doi:10.1016/j.appdev.2013.10.004.

Gasser, L., & Keller, M. (2009). Are the competent the morally good? Perspective taking and moral motivation of children involved in bullying. *Social Development*, *18*, 798–816. doi:10.1111/j.1467-9507.2008.00516.x.

Gini, G. (2006). Social cognition and moral cognition in bullying: What's wrong? *Aggressive Behavior*, *32*, 528–539. doi:10.1002/ab.20153.

Hardy, S. A., Bhattacharjee, A., Reed, A. II, & Aquino, K. (2010). Moral identity and psychological distance: The case of adolescent parental socialization. *Journal of Adolescence 33*, 111–123. doi: 10.1016/j.adolescence.2009.04.008

Hardy, S. A., & Carlo, G. (2011). Moral identity: What is it, how does it develop, and is it linked to moral action? *Child Development Perspectives*, *5*, 212–218. doi:10.1111/j.1750-8606.2011.00189.x.

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review*, *102*, 458–489. doi:10.1037/0033-295X.102.3.458.

Harter, S. (1985). *Manual for the Self-perception Profile for Children.* Denver, CO: University of Denver.

Harter, S. (1999). *The construction of the self: A developmental perspective.* New York, NY: Guilford.

Hawkins, D. L., Pepler, D. J., & Craig, W. M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development*, *10*, 512–527, doi:10.1111/1467-9507.00178

Hughes, C., & Leekam, S. R. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development*, *13*, 590–619. doi:10.1111/j.1467-9507.2004.00285.x.

Jackson, M., & Tisak, M. S. (2001). Is prosocial behavior a good thing? Developmental change in children's evaluations of helping, sharing, cooperating, and comforting. *British Journal of Developmental Psychology, 19*, 349–367. doi:10.1348/026151001166146.

Ji, P., DuBois, D. L., & Flay, B. R. (2013). Social emotional and character development scale: Development and initial validation with urban elementary school students. *Journal of Research in Character Development*, *9*, 121–147.

Keller, M. (1984). Resolving conflicts in friendships: The development of moral understanding in everyday life. In W. M. Kurtines, & J. L. Gewirtz (Eds.), *Morality, moral behavior, and moral development* (pp, 140–158). New York, NY: John Wiley & Sons.

Krettenauer, T., & Eichler, D. (2006). Adolescents' self-attributed moral emotions following a moral transgression: Relations with delinquency, confidence in moral judgment and age. *British Journal of Developmental Psychology*, *24*, 489–506. doi:10.1348/02615005X50825.

Krettenauer, T., Malti, T., & Sokol, B. (2008). The development of moral emotion expectancies and the happy victimizer phenomenon: A critical review and application. *European Journal of Developmental Science*, *2*, 221–235.

Kyratzis, A. (2004). Talk and interaction among children and the co-construction of peer groups and peer culture. *Annual Review of Anthropology*, *33*, 625–649. doi:10.1146/annurev.anthro.33.070203.144008.

Ladd, G. W., & Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive-withdrawn children during early grade school. *Child Development*, *70*, 910–929. doi:10.1111/1467-8624.00066

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development*, *67*, 1103–1118. doi:10.2307/1131882

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, *4*, 231–244. doi:10.1556/JCEP.4.2006.3-4.3.

Mant, C. M., & Perner, J. (1988). The child's understanding of commitment. *Developmental Psychology*, *24*, 343–351.

Murray-Close, D., Crick, N. R., & Galotti, K. M. (2006). Children's moral reasoning regarding physical and relational aggression. *Social Development*, *15*, 345–372. doi:10.1111/j.1467.2006.00346.x.

Nucci, L., & Smetana, J. G., (1996). Mothers' conceptions of young children's areas of personal freedom. *Child Development*, *67*, 1870–1886. doi:10.2307/1131737.

Nucci, L., & Weber, E. K. (1995). Social interactions in the home and the development of young children's conceptions of the personal. *Child Development*, *66*, 1438–1452. doi:10.2307/1131656.

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development*, *59*, 1323–1338. doi:10.2307/11304049.

Park, Y., & Killen, M. (2010). When is peer rejection justifiable? Children's understanding across two cultures. *Cognitive Development*, *25*, 290–301. doi:10.1016/j.cogdev.2009.10.004.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin*, *102*, 357–389. doi:10.1027/0033-2909.102.3.357.

Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology*, *29*, 611–621. doi:10.1037/0012-1649.29.4.611.

Peterson, C. C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology*, *20*, 205–224. doi:10.1348/026151002166415.

Peeters, M., Cillessen, A. H. N., & Scholte, R. H. J. (2010). Clueless or powerful: Identifying subtypes of bullies in adolescence. *Journal of Youth and Adolescence*, *39*, 1041–1052. doi:10.1007/s10964-009-9478-9.

Piaget, J. (1932/1965). *The moral judgment of the child*. New York, NY: Free Press.

Reijntjes, A., Vermande, M., Olthof, T., Goossens, F. A., van de Schoot, R., Aleva, L., & van der Meulen, M. (2013). Costs and benefits of bullying in the context of the peer group: A three-wave longitudinal analysis. *Journal of Abnormal Child Psychology*, *41*, 1217–1229. doi:10.1007/s10802-013-9759-3.

Recchia, H., Wainryb, C., & Pasupathi, M. (2013). 'Two for flinching': Children's and adolescents' narrative accounts of harming their friends and siblings. *Child Development*, *84*, 1459–1474. doi:10.1111/cdev.12059.

Rodkin, P. C., & Roisman, G. I. (2010). Antecedents and correlates of the popular-aggressive phenomenon in elementary school. *Child Development*, *81*, 837–850. doi:10.1111/j.1467-8624.2010.01437.x.

Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, *15*, 112–120. doi:10.1016/j.avb.2009.08.007.

Sandstrom. M., Makover, H., & Bartini, M. (2013). Social context of bullying: Do misperceptions of group norms influence children's responses to witnessed episodes? *Social Influence*, *8*, 196–215. doi:10.1080/15534510.2011.651302.

Shaw, L. A., & Wainryb, C. (2006). When victims don't cry: Children's understanding of victimization, compliance and subversion. *Child Development*, *77*, 1050–1062. doi:10.1111/j.1467-8624.00918.x.

Shi, B., & Xie, H. (2012). Popular and nonpopular subtypes of physically attractive preadolescents: Continuity of aggression and peer mechanisms during the transition to middle school. *Merrill-Palmer Quarterly*, *58*, 530–553.

Siegal, M. (1982). *Fairness in children: A social-cognitive approach to the study of moral development.* London: Academic Press.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.) *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Lawrence Erlbaum.

Smetana, J. G., Killen, M., & Turiel, E. (1991). Children's reasoning about interpersonal and moral conflicts. *Child Development*, *62*, 629–644. doi:10.2307/1131136.

Smetana, J. G., Kochanska, G., & Chuang, S. (2000). Mothers' conceptions of everyday rules for young toddlers: A longitudinal investigation. *Merrill-Palmer Quarterly*, *46*, 391–416.

Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., & Comer, J. (2012). Developmental changes and individual differences in young children's moral judgments. *Child Development*, *83*, 683–696. doi:10.1111/j.1467-8624.2011.01714.x.

Sutton, J. (2003). ToM goes to school: Social cognition and social values in bullying. In B. Repacholi & V. Slaughter (Eds.). *Individual differences in theory of mind: implications for typical and atypical development* (pp. 99–120). New York: Psychology Press.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology*, *17*, 435–450. doi:10.1348/026151099165384.

Troop-Gordon, W., Visconti, K. J., & Kuntz, K. J. (2011). Perceived popularity during early adolescence: Links to declining school adjustment among aggressive youth. *Journal of Early Adolescence*, *31*, 125–151. doi:10.1177/0272431610384488.

Wainryb, C., Brehl, B. A., & Matwin, S. (2005). *Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal conflicts.* Monographs of the Society for Research in Child Development. Oxford: Blackwell Publishing.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684. doi:10.1111/1467-8624.00304.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. doi: 10.1016/0010-0277(83)90004-5.

22 }

Character across Early Emerging Adulthood: Character Traits, Character Strivings, and Moral Self-Attributes

Erik E. Noftle

This chapter describes a two-year longitudinal study of early emerging adulthood, which focused on three different aspects of character: self-views of moral behavioral standing (*moral self-attributes*), trait-like behavioral tendencies (*character traits*), and chronically activated motivations (*character strivings*). The purpose of the chapter is threefold: (1) to examine what character is like at the beginning of emerging adulthood—the years immediately after the teenage years, (2) to gauge the usefulness of studying explicit self-views of morality in character research, and (3) to identify *character prospection* as a missing piece in the study of character, and demonstrate its potential contribution to understanding character.

The chapter has the following four-part structure. First, a proposal is made that emerging adulthood is an essential period in which to study character, because it is a transformative period of self-development. Second, the study of character in personality psychology is briefly reviewed and three different aspects of character are presented in detail: the aforementioned character traits and character strivings and a new aspect, that of moral self-attributes. Third, the current research is presented, which pursues three research questions: (1) How does character stand at the beginning of emerging adulthood—which aspects of character are typically descriptive of emerging adults, and how do these change across the first two years of college? (2) How are the three different aspects of character connected—how tightly or loosely do moral self-attributes connect with behavioral and motivational aspects of character? and 3) What do character prospection and related phenomena in this period look like—which character strivings do emerging adults find highly important, what are their ideal levels for character traits and moral self-attributes, and which of these do

they actually plan to improve upon and change? Finally, future directions are suggested for the study of character in emerging adulthood.

Why Study Character in Emerging Adulthood?

The current research questions explore the developmental period of emerging adulthood as a transitional phase between adolescence and adulthood (Arnett, 2000). There are at least three reasons why emerging adulthood is a particularly important period in which to study character development: (1) particular features of the developmental period itself, (2) suggestions from moral development theory about the emerging adulthood period, and (3) previous evidence establishing a sizable, pervasive pattern of normative personality trait development during the period.

First, emerging adulthood demonstrates unique features as a developmental period. Arnett (2000) argued that across many contemporary cultures, especially those of industrialized or postindustrial countries in North America and Europe (although not limited to them), young people in their late teens and early twenties occupy a distinct developmental period that can be characterized neither as adolescence nor as full-fledged adulthood. Individuals in this age span of 18 to 25 (or 30) years tend to have gained some measure of independence from their parents yet generally have not adopted conventionally adult roles within marriage, parenthood, and stable careers. Arnett describes this period as the most unstable period of life—constituting instability in relationships, living circumstances, jobs, and identity. However, more positively, it is also posited as the most "volitional" stage in life (Arnett, 2000; p. 469). It is a stage during which new freedoms, responsibilities, and identities are taken on, and individuals are faced with questions of whom and—most germane to the current study—*how* they want to be.

Second, moral development theory suggests that emerging adulthood is a critical formative period for moral maturity in adulthood. One early personality theory of moral development was formulated by Robert Hogan and his colleagues, who were some of the first to argue against the cognitive developmental zeitgeist dominating moral psychology from the 1950s through at least the 1980s. Hogan and his colleagues were dismayed by what they saw as a narrow focus on moral reasoning and argued that, instead, moral development is accompanied by developmental periods that foster broader personality characteristics. Specifically, they argued that these three central personality characteristics, socialization (or rule attunement), empathy (or social sensitivity), and autonomy (or independence and self-awareness in moral action) are first learned and then internalized (Hogan, 1973; Hogan, Johnson, & Emler, 1978). Although socialization and empathy are usually gained by the mid-teenage years, Hogan argued that growing autonomy is a hallmark of

what is now identified as the emerging adult period.[1] Current perspectives reaffirm the importance of studying moral development past the period of adolescence (Hill & Roberts, 2010).

Third, another reason to focus on emerging adulthood to understand character development is research demonstrating robust patterns of personality trait development during the period. This literature not only reveals that emerging adulthood is typically a dynamic period of development, but also demonstrates specific personality changes that relate closely to character (a point that will be discussed at more length in the next section). Roberts, Walton, and Viechtbauer (2006) conducted a meta-analysis of longitudinal studies of personality change, revealing that the period of most sizable, and most positive, personality change was emerging adulthood, in which statistically significant increases were identified for four of the Big Five traits: extraversion, conscientiousness, emotional stability, and openness (a non-significant trend for increases in the fifth trait, agreeableness, was also noted during the period). Significant change has even been found over shorter spans of time within college (Asendorpf & Wilpers, 1998; Davis & Franzoi, 1991; Plant & Telford, 1966; Vaidya, Gray, Haig, & Watson, 2002), and studies reveal greater trait change across the college years than immediately after (e.g., Vaidya, Gray, Haig, Mroczek, & Watson, 2008). These findings suggest that early emerging adulthood—the traditional college years—could represent an ideal time to study changes in character, which may parallel and be similarly dynamic to the trait changes that have already been observed.

Although the basic patterns of trait changes by themselves suggest that emerging adulthood could be an ideal period in which to study character development, studying trait changes is not enough. It also is important to examine emerging adults' volitional motivations and aspirations to change, which may in part underlie changes that actually take place across the emerging adult period. This emphasis on *character prospection*, or how people project their character to change in the near future, contributes to a growing interest in prospection and self-regulation of development in personality (Haase, Heckhausen, & Wrosch, 2013; Hennecke, Bleidorn, Denissen, & Wood, 2014; Noftle, 2010, and in preparation), as well as more generally in psychology (Seligman, Railton, Baumeister, & Sripada, 2013). The current research asks about which character strivings are particularly important to emerging adults to find out what aspects of character they are particularly motivated to pursue. In addition, participants are asked to contrast their current levels of character traits and moral self-attributes both to their predicted levels by the end of college and their ideal levels of those characteristics. Thus, this uncovers which

[1]Studies of moral reasoning also demonstrate that it develops considerably during young adulthood (e.g., Armon & Dawson, 1997).

characteristics they ideally want to improve, and the changes in character they expect to occur across early emerging adulthood. Although a few studies have inquired about emerging adults' prospected personalities, so far they are few in number and have all focused on broad traits and relatively long time periods (Fleeson & Heckhausen, 1997; Haslam, Bastian, Fox, & Whelan, 2007; Krueger & Heckhausen, 1993; Wood & Roberts, 2006).

Arriving at a deeper understanding of how character forms and is expected to grow during this period will reveal how moral development may be fostered and achieved, and possibly suggest specific areas of typical character neglect. Examining the character of emerging adults is especially timely in the current era, in which young people have been argued to be excessively self-focused and egocentric, a moral failing held to be uncharacteristic of previous generations (Twenge & Campbell, 2009; but see Arnett, 2010; Trzesniewski, Donnellan, & Robins, 2008).

What Is Character? A Personality Perspective

Research into character can be pursued in at least two ways: by studying cognitive processes and by studying the positive characteristics of individuals.[2] Until recently, the dominant tradition in moral psychology took a cognitive approach to studying moral character through moral reasoning. This cognitive approach tended to examine how people process moral quandaries and the types of resulting decisions they make. The cognitive approach also dismissed an Aristotelian "bag of virtues" approach (Kohlberg, 1981, p. 2, as cited in Cawley, Martin, & Johnson, 2000) that focused on fostering a plethora of dispositional characteristics such as honesty, generosity, courage, etc. (Walker & Frimer, 2009). On the other hand, an alternative approach to studying character is deeply rooted in personality psychology's study of such characteristics, including personality traits (Noftle, Schnitker, & Robins, 2011).

Not too long ago a personality trait-like approach to character was deemed to be essentially moribund in moral psychology (Lapsley, 1996, p. 213). This judgment possibly stems from two influences. First, there was an early suspicion against studying moral character within personality, which was espoused by the influential personality psychologist Gordon Allport in the 1930s (McCullough & Snyder, 2000; Nicholson, 1998). Allport argued for moving personality away from the study of moral characteristics definable in terms of "conventional meanings" that were historically contextualized, and thus less

[2] Certainly a third valuable pursuit for research into character is negative personality characteristics (e.g., the Dark Triad; Paulhus & Williams, 2002), but the current approach focuses on positive characteristics that can be seen as representative of virtues (rather than vices).

scientific (Allport, 1931). In their pioneering studies of traits in the lexicon, Allport and Odbert (1936) excluded from consideration highly evaluative terms (e.g., "average," "superior," "worthy," "likeable"), including many having to do with moral character ("ethical," "honorable," "moral," etc.), to focus instead on description. The lexical approach to personality, which traces its roots to Allport and Odbert's (1936) seminal work, continues to strongly influence the study of personality (Goldberg, 1993; John & Robins, 1993a). Allport's (1937) conclusion, stated in his influential personality textbook, was that "[c]haracter is personality evaluated, and personality is character devaluated" (p. 52). Allport clearly believed character shouldn't be the focus of a scientific pursuit of personality, and many agreed with him (but see Noftle et al., 2011).

Second, there was a crisis within personality psychology itself. The highly influential results of the Hartshorne and May studies of honesty and other character traits (Hartshorne, May, Maller, & Shuttleworth, 1928), seemingly showing little coherence or consistency in such traits, dampened enthusiasm for research pursuits in moral character (Cawley et al., 2000; Hogan, 1973). In addition, the critiques of trait psychology by Mischel (1968) and the effects of the person-situation debate that followed (Fleeson & Noftle, 2008; Miller, 2014) slowed research on traits more broadly. However, research on traits has since rebounded (Funder, 1991), particularly within the study of the Big Five, which have received increasing attention since the mid-1990s (John, Naumann, & Soto, 2008). Moreover, a personality approach to studying character has returned to moral psychology and is beginning to gain broader attention (Hill & Roberts, 2010). This approach forms the focus of the current research. Two existing personality trait taxonomies are considered that capture a lot about character: the Big Five broad traits and facets and the Values in Action (VIA) character strengths.

ONE APPROACH TO CHARACTER: THE BIG FIVE PERSONALITY TRAITS AND FACETS

The Big Five is a hierarchical taxonomy of traits, which attempts to organize all of the trait-relevant ways in which people differ from one another. At the basic level it includes the five relatively independent, broad traits of extraversion (e.g., talkative, sociable, assertive), agreeableness (kind, cooperative, altruistic), conscientiousness (organized, responsible, industrious), emotional stability (calm, even-tempered, imperturbable), and openness to experience (creative, intellectual, reflective). At a lower subcomponent "facet" level, the Big Five contains sets of narrower, more specific traits that are subsumed under one of the five broader traits (e.g., Costa & McCrae, 1992; Saucier & Ostendorf, 1999). The traits included within the Big Five taxonomy have their origins all the way back in Allport and Odbert's (1936) list of trait terms (John, et al., 2008). However, establishment of the Big Five as a consensually accepted taxonomy

of traits only emerged in the early 1990s (Goldberg, 1993; McCrae & John, 1992). After decades of research, there is now a great deal of support for the cross-cultural universality, heritability, judgability, temporal stability, and predictive validity of the Big Five (John et al., 2008).

Although the Big Five includes content that is descriptive of a person's personality traits, it clearly also includes evaluative content that is related to character. The first piece of evidence is that the traits of extraversion, agreeableness, conscientiousness, emotional stability, and openness are all socially desirable, albeit to varying degrees. Although certainly not all that is socially desirable is relevant to good character, it would be surprising if people with good character were not more likely to be living up to the standards of a typical society. Supporting the social desirability of the Big Five, individuals with higher self-esteem tend to be higher on all of the Big Five dimensions, including emotional stability, extraversion, and conscientiousness (Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001). According to the sociometer theory of self-esteem (Leary, Tambor, Terdal, & Downs, 1995), this suggests that individuals with higher levels of these traits believe they are accepted by and living up to the standards of society. Another literature which suggests the broad social desirability across the Big Five is that of subjective well-being, or how people evaluate their lives. People who tend to be higher on the Big Five traits tend to report greater well-being (DeNeve & Cooper, 1998; Schimmack, Oishi, Furr, & Funder, 2004), and greater well-being also suggests better adjustment to society.

A second piece of evidence that links the Big Five to character has to do with the beneficial outcomes of being higher on the Big Five traits. This means that personality is directly evaluated in terms of its actual, real-world consequences (something Allport might have deemed too evaluative). As with social desirability, it would be surprising if people with good character were not more likely to be flourishing, at least in some domains. Although "niche" models having some adaptive or beneficial qualities have been proposed for the low pole of each dimension (Nettle, 2006), so far, most of the findings concerned with Big Five prediction suggest that it is better on average to be higher on the traits than lower on them. Ozer and Benet-Martinez (2006) conducted a literature review of real-world outcomes of the Big Five and found that higher levels of the five traits were related to better outcomes at the individual, interpersonal, and societal levels.

A third, perhaps even more important, piece of evidence stems from a study that sought to examine virtue as an aspect of personality and employed the lexical approach to create a virtues scale (Cawley et al., 2000). Cawley and his colleagues searched the dictionary for words that displayed qualities that one ought to embody or display (i.e., virtues), wrote items related to those qualities, and then factor-analyzed the resulting set of 140 items. They identified four very broad virtue factors: empathy, order, resourcefulness, and serenity.

Interestingly, these broad virtue dimensions were associated moderately to strongly with the Big Five. For example in their first larger study, correlations of |.40| (absolute value of .40) or above were found between the virtue factors and all of the Big Five except openness. McCrae and John (1992) once suggested that agreeableness and conscientiousness represent "the classic dimensions of character" and specifically have "moral overtones" (p. 197). However, such sizable correlations suggest that character extends beyond just those two broad traits, and that at least four of the Big Five might capture character.

However, importantly, Cawley and colleagues also showed that the correlations between the four virtues and the Big Five varied considerably at the subcomponent, or facet, level. For example, empathy correlated .64 with the warmth facet of extraversion, correlated .48 with the positive-emotions facet, but only correlated .12 with the excitement-seeking facet. As discussed previously, openness didn't have a strong relation at the broad trait level with any of the virtue factors; however, at the facet level, openness to feelings correlated .50 with empathy, which is not only a sizable but also clearly a psychologically meaningful association. These findings suggest at least three things. First, they suggest that the broad Big Five level of analysis might conflate character traits with qualities that might be psychologically co-occurent but not related to character per se. Perhaps this means psychologically that people with certain character traits are also more likely to have higher levels of certain traits that are unrelated to character. Second, the findings suggest that Cawley's broad virtue dimensions capture a lot but might be too broad. For example, empathy is highly correlated with both warmth and openness to feelings (which are not as strongly correlated with one another); perhaps both of those Big Five facets represent distinct character qualities that both facilitate being empathic. Third, the findings affirm abandoning attempts to isolate a singular moral personality (Hill & Roberts, 2010), and instead would argue that the Aristotelian bag-of-virtues approach might actually be worthwhile. In particular, it suggests that a bag of virtues stuffed with Big Five facets, and thus pursuing a facet level of analysis rather than a broad level of analysis, might be quite useful for studying character.

In retrospect, such close associations between broad virtues and the Big Five are perhaps not surprising. In general, people consider someone who is friendly, compassionate, responsible, and tolerant—higher on Big Five traits—to have good character (Lapsley & Lasky, 2001). In fact, McAdams (2009) has recently argued that these traits "carry considerable moral meaning" (p. 14). Thus, the current study considers character traits as a *subset* of the Big Five personality traits, studied at a *narrower level* of analysis, at the facet level rather than the broad domain level. First, the notion of a subset is important because, certainly, not all Big Five traits are indicative of good (or bad) character (e.g., anxiety). Second, the notion of a *narrower level* of analysis is important, because the Big Five considered at the broadest level may actually

combine traits that reflect good character, traits that are unrelated to character, and perhaps even traits that suggest bad character. For example, extraversion includes assertiveness, which reflects character traits such as leadership, but also includes excitement-seeking, which is not generally held to be a trait of good character. Thus, it is important to examine Big Five facets, rather than just the broad traits, to gain insight into character.

Although the Big Five is a useful taxonomy for studying character, it does have potential downsides. It draws from the universe of the lexicon, which means it should be all-inclusive. However, the Big Five was also developed to be parsimonious—to assess the main ways in which people's psychological traits differ, not every single possible difference—and has sometimes been referred to as an "adequate taxonomy" (Norman, 1963). In addition, it both includes traits that are irrelevant to good character and likely leaves out some important traits that have to do with character. Indeed, there are some identifiable traits that have to do with character that have been demonstrated to mostly lie outside the domain of the Big Five (e.g., broad traits, such as honesty-humility [Ashton & Lee, 2005]; and narrow traits like thriftiness [Paunonen & Jackson, 2000]).[3]

Next is considered another much more recently developed trait taxonomy that may be useful in studying character.

AN ALTERNATIVE APPROACH TO CHARACTER: THE VIA CHARACTER STRENGTHS

In the last decade, a different approach to studying character has been pursued, one apart from the lexical approach of the Big Five but one that still examines character as traits: the VIA character strengths (Peterson & Seligman, 2004). Developed to serve as a "manual of the sanities," or a classification of mental health and flourishing comparable to the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994), the VIA seeks to delineate individual differences in the complex nuances of character. The VIA comprises a taxonomy of 24 character strengths organized under six umbrella virtues: wisdom, courage, humanity, justice, temperance, and transcendence. The set of 24 character strengths were identified theoretically from an impressive, seemingly near-comprehensive survey of world religious, social, and intellectual traditions (Peterson & Seligman, 2004) that spanned at least three years (McGrath, 2014). In this classification system, the broader virtues are defined

[3] Honesty-humility was initially considered to be included in the current empirical study as one of the broader traits, alongside the Big Five. However, little work developing and validating adjectival honesty-humility scales had been conducted at the time the study was begun. However, sincerity was included as one of the four facets of agreeableness, which clearly reflects honesty-humility content but conforms to agreeableness as it is assessed in the Five-Factor model (Ashton & Lee, 2005).

as the core characteristics valued by moral philosophers and religious thinkers across time and cultures. These virtues are construed as broad universals, rooted in biology, that have been evolutionarily selected. Character strengths fall under these "high six" broad virtues and are conceptualized as the psychological processes and mechanisms by which virtue may be achieved. For example, a person may display the virtue of transcendence through a variety of routes: gratitude, hope, humor, spirituality, or appreciation of beauty, five of the character strengths. Although the hierarchical structure of virtues containing sets of character strengths is theoretically interesting, it hasn't been borne out empirically in factor or cluster analyses (McGrath, 2014; Noftle, Schnitker, & Robins, 2011). At any rate, most research tends to focus on the character strengths rather than the broader virtues.

A measure, the VIA-IS, has been developed to assess each of the 24 different strengths and has received a lot of use. In the Values in Action Inventory of Strengths (VIA-IS), character strengths are measured in the same way traits are sometimes measured (such as in the NEO Personality Inventory–Revised [NEO-PI-R]), as statements about what a person is like in general, which require either agreement or disagreement. Thus, both the conception of character strengths and the measurement strategy lead them to be classified into the trait domain within personality alongside the Big Five. One advantage of the character strengths over the Big Five is that they were developed to specifically assess character traits rather than traits in general. One downside is that the rational method used in developing the taxonomy may not guarantee full inclusiveness of possible character content. For example, patience has been raised as a commonly recognized character strength across cultures but is not included in the VIA (Schnitker & Emmons, 2007). In addition, the VIA only assesses components of good character; it does not assess bad character (Peterson & Seligman, 2004).

Connections between the Big Five and VIA Character Strengths

Although the two taxonomies were developed separately, using very different methods, several studies reveal a close connection between Big Five traits and character strengths, demonstrating a substantial amount of variance in character strengths that can be explained by the Big Five; in other words, there is a sizable overlap between the two (De Raad & Van Oudenhoven, 2011; MacDonald, Bore, & Munro, 2008; Noftle, Schnitker, & Robins, 2011; Steger, Hicks, Kashdan, Krueger, & Bouchard, 2007). For example, Noftle et al. (2011) assessed the relationship between a rather comprehensive Big Five measure, the NEO-PI-R (Costa & McCrae, 1992), which includes six narrower facet scales for each broad trait, and the VIA character strengths (Peterson & Seligman, 2004). For each of the 24 VIA character strengths, the researchers identified the individual Big Five facet that was most highly positively correlated with it, and demonstrated that the average correlation between these was $r = .47$

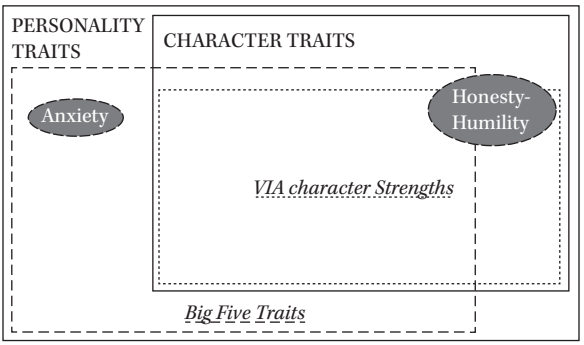

FIGURE 22.1 Conceptual Mapping of the Big Five and Values in Action (VIA) Character Strength Taxonomies into Personality Trait and Character Trait Conceptual Space.

(.65 corrected for attenuation), displaying considerable overlap. The resulting 14 Big Five facets represented were numerous and diverse, reflecting the fact that character strengths were moderately to strongly associated with a variety of narrower traits within the Big Five: extraversion (4 facets), agreeableness (4 facets), conscientiousness (4 facets), and openness (2 facets).[4] Only the VIA strength of spirituality was just *moderately* associated ($r = .36$) with its closest facet after correcting for attenuation; all other strengths were correlated at .50 or greater. Although such strong associations suggest a substantial amount of shared content between the two taxonomies, they do not suggest complete redundancy. One conclusion that can be made from this research is that the Big Five facets do actually include many aspects of character. Another conclusion is that character strengths do seem to capture at least some aspects of character not included in the Big Five.

Figure 22.1 maps the hypothesized relations among the conceptual universes of personality traits and character traits and the two taxonomic measurement systems of the Big Five traits and the VIA character strengths. The largest category in figure 22.1 is personality traits, which is considered to completely subsume the category of character traits. As previously discussed, not all personality traits are relevant or indicative of character, but all character traits are seen as being within the universe of personality traits. The Big Five traits make up a substantial portion of the universe of personality traits, but not all of it. The VIA character strengths make up a substantial portion of the universe of character traits, but not all of it. Importantly, the Big Five traits and VIA character strengths overlap a considerable amount, as demonstrated by

[4] If neuroticism was reversed as emotional stability, only one of the 24 character strengths had its highest positive correlation with it. As has been found previously, character is more weakly related to emotional stability than are the other four traits (e.g., De Raad & Van Oudenhoven, 2011; Peterson & Seligman, 2004); but see Cawley et al. (2000) for a model that includes more content related to emotional stability.

the research discussed in the previous paragraph. There are Big Five facet-level traits which are not considered character traits, such as anxiety. There are broad character traits, such as honesty-humility, which overlap with the Big Five a little bit, but are more outside the Big Five than within it (Ashton & Lee, 2005). It is expected that there is probably some Big Five content that is pertinent to character traits which is not covered by the VIA character strengths but that there is likely more VIA character strength content that is relevant to character traits that is not covered by the Big Five.

A RECONCEPTUALIZATION OF CHARACTER STRENGTHS: REFORMULATING CHARACTER STRENGTHS AS CHARACTER STRIVINGS

Because of the considerable overlap between the Big Five and VIA character strengths, rather than simply measure both types of traits, which would lead to a lot of redundancy, character strengths were reimagined to reflect a different, motivational, type of construct in personality. When one considers what it means to have good character, it can be an endpoint (e.g., having a good character trait); however, it also can represent a recurring motive—a self-motivated aim instead of a fairly static quality. For example, gratitude can describe a person (e.g., a grateful person), but it also can describe a person's motivation (e.g., to typically try to appreciate every day). In fact, virtues and character strengths are often proscriptive (in addition to being descriptive); in religious, philosophical, and cultural teachings, virtues are often held as exemplary ideals which should be strived toward constantly. In this way, character can be partially measured as a different part of personality, that of motivated strivings, which are characteristic goals chronically pursued by individuals (Emmons, 1986).[5] Although this type of goal construct has received a lot of study, it has only begun to be integrated into the study of character and virtue (e.g., Emmons, 2003).

The current research attempts to further advance the study of strivings in character. In fact, Peterson and Seligman's (2004) character strengths were originally conceptualized as reflecting processes or mechanisms which bring about virtue, which would suggest they should in fact have a dynamic quality. Thus, character strengths can easily be reimagined as *character strivings* by rephrasing the content in terms of what a person is typically trying to achieve, thus incorporating the same essential content from the character strength taxonomy. Because the character strengths in the VIA-IS consist of short

[5] Strivings contribute to a larger category of characteristic adaptations in personality—characteristics that can be conceptualized as more dynamic, contextualized, and oriented toward goals (e.g., Hooker, 2002). Indeed, the two categories of traits and of characteristic adaptations are often considered to represent the two broadest types of constructs in personality (Cantor, 1990).

statements, they can easily be altered and rephrased to capture strivings rather than strengths.[6] In this way, character strivings become aspirational and motivational rather than only descriptive, as they stood in the form of character strengths.[7] Studying character strivings is important because it represents one way of identifying the specific virtues emerging adults are motivated to achieve and suggests one social-cognitive mechanism through which such gains might actually be achieved.[8] Studying character strivings together with character traits advances the understanding of the connections between goals and traits, which previously have been largely pursued independently (Hooker & McAdams, 2003; although see Walker & Frimer, 2009). More broadly, it also takes a step toward bringing together trait and social-cognitive approaches to character which have only rarely been pursued in unison (Lapsley & Hill, 2009).

A NEW CONCEPTUALIZATION FOR CHARACTER: MORAL SELF-ATTRIBUTES

The last of the three aspects of character, after character traits and character strivings, represents a new approach to studying character: moral self-attributes. Although many would say that certain character traits and character strivings are conceptually strongly related to morality (e.g., altruism, integrity, fairness), other named aspects of character that are socially esteemed would appear to be unrelated to morality (e.g., assertiveness, creativity, zest). But do character traits and character strivings together completely map out what it means to see oneself as moral—surely an important part of character? Three studies on moral identity suggest not. Walker and Pitts (1998) conducted a set of studies on lay conceptions of morality, asking people for terms and phrases that describe moral exemplars. A sizable percentage of the terms that participants generated had to do with agreeableness and conscientiousness, as well as related character strengths such as fairness, prudence, self-control, and honesty. However, there was a cluster of more evaluative terms that the

[6] Certainly, the decision to reformulate character strengths into strivings could have alternatively been done with Big Five (e.g., I typically try to be compassionate). However, it was thought that the additional context inherent in the character strength statements would better lend itself to formulating strivings.

[7] It should be noted that the character strivings that are highlighted in the current study are different from Emmons's (1986) original conception of personal strivings in a crucial way: character strivings are nomothetic (e.g., content generated and provided *by* researchers that is rated by participants) rather than idiographic like personal strivings (e.g., content generated and provided *to* researchers by participants themselves).

[8] One important question is what the true object of the character striving is. For example, if a person wants to be honest more of the time, is it for the ideal of being an honest person (i.e., a virtuous end), or is it for some other purpose, such as being more authentic, reducing anxiety, avoiding being caught in a lie, or something else (i.e., where the virtue is the means for some other end)? This distinction will be vital to examine in future research.

researchers termed *principled-idealistic* that included content not present in the wide-ranging sets of character traits and strivings. This cluster was explicitly about being a moral person and included such terms as *ethical* and *principled*, as well as a number of phrases about right versus wrong, convictions, and conscience. Similarly, Lapsley and Lasky (2001) found a preponderance of agreeableness and conscientiousness terms among the most prototypic terms used to describe a virtuous person who has good character. However, like Walker and Pitts, they also found content specifically about being a good person: terms like *moral, honorable*, and *ethical* were in the top third of prototypicality ratings. A third study, by Saucier and Goldberg (1998), found that self-ratings of morality (including such terms as *moral* and *ethical*) were only correlated modestly, under .30, with any of the Big Five (see also Paunonen & Jackson, 2000). However, to my knowledge, research has not examined self-ratings of morality further.[9]

The current research identified a set of six terms similar to the ones previously mentioned to assess this explicitly evaluative moral domain: "moral," "ethical," "principled," "virtuous," "honorable," and "a person of good character." Unlike the studies of moral exemplars, the current research on moral self-attributes places the focus specifically on *self-views* of morality as a specific part of character, which emphasize the individual's own moral code and how well they live up to it. In particular, the person is asked to think explicitly about how much they typically behave morally, ethically, and so on.[10] Thus, moral self-attributes differ in some major ways from character traits and strivings, but likely share some similarities with moral identity, as seen in the content considered in the previous paragraph. Table 22.1 outlines some major conceptual differences among the three aspects of character as measured in the current study, highlighting how moral self-attributes are different. The first three are held to be definitional to the constructs themselves. The latter three are more dependent on how they are chosen to be assessed.

First, as mentioned previously, moral self-attributes are conceptualized as an aspect of character that is narrowly concerned with morality per se, unlike the broader sets of traits and strivings. What is revealed is the individual's own judgment of how morally they typically behave, not any specific character qualities that contribute to that judgment, qualities that could vary across individuals. Second, moral self-attributes are conceptualized as constituting a singular

[9] Paunonen created a measure, the Supernumerary Personality Inventory, to assess the 10 clusters of psychological attributes that didn't overlap strongly with the Big Five, of which morality/honesty was one, but focused the resulting scale mostly on dishonesty and immorality content rather than on morality content.

[10] Although such terms as *moral* and *ethical* connote quite different meanings in scholarly work, they tend to be rated very similarly as self-descriptors by laypeople, and thus are combined into a single scale.

TABLE 22.1 } **Conceptual Differences among the Moral Self-Attributes, Character Traits, and Character Strivings Assessed in the Current Research**

Character attribute	Moral self-attributes	Character traits	Character strivings
Explicit concern with morality and ethics	High	Low	Moderate
Amount of content coverage	Low	High	High
Subjectivity	High	Low	Moderate
Specificity in terms of behaviors	Low	Moderate	High
Contextualization (in the current measures)	Low	Low	Moderate
Connection with constructs typically currently studied in personality psychology	Low	High	High

evaluation. As such, adequate content coverage of moral self-attributes in assessment is much more easily obtainable; only a single scale is required instead of the couple dozen needed for a reasonably comprehensive assessment of either character traits or character strivings. Third, the subjectivity of the ratings of moral self-attributes is considered to be highest among the three character aspects. Although Allport was concerned that including such evaluative content in personality was problematic because of the changing meanings of such highly socially constructed content over time, there is also a strong benefit to assessing how the individual subjectively views him or herself. On the other hand, the subjectivity of character traits is low, or at least considerably lower; for example, there's considerable agreement between self-reports and informant reports (reports by well-acquainted others) for most traits (John & Robins, 1993b; McCrae, 1982; Watson, Hubbard, & Wiese, 2000). Findings that agreement between self-reports and informant reports is similarly high for character strengths (Ruch et al. 2010) suggests the same might be true for character strivings. Fourth, the specificity in behavior of moral self-attributes is low; rating oneself as typically acting morally or not morally, or acting honorably or not honorably, because they are both broad and abstract qualities, does not provide a clear indication of specifically what a person might do.[11] On the other hand, ratings of Big Five adjectives that assess character traits (e.g., considerate) have a closer link to specific behaviors. Character strivings, which include phrases, present content with an even closer connection to behavior (e.g., "engage myself in many different activities").[12] Fifth, moral self-attributes are completely decontextualized; no indication of what particular kinds of situations are involved is included. Although this makes moral self-attributes

[11] Although not all of the six moral terms are exactly equal in specificity in terms of behavior (e.g., *moral* is probably broader than *principled*), they tend to have much lower specificity than the character trait terms included (e.g., talkative).

[12] It is worth noting that both the Big Five and character strivings could have either moderate or high specificity in behavior depending on how they are assessed.

similar to character traits, because they both rely on adjectives, character strivings are quite different because their phrase-like nature includes more contextualization (e.g., "respect others' opinions in a group that I'm in" vs. "respectful").[13] Sixth, unlike character traits and character strivings, constructs that connect strongly to current topics studied by personality psychologists, moral self-attributes are currently more novel. An advantage of the current work is that it includes these three very different aspects of character, which avoids previously highlighted drawbacks of attempting to identify a singular moral personality (Hill & Roberts, 2010).

Current Research

The current research examined three aspects of character: character traits, character strivings, and moral self-attributes, and pursued three questions: (1) What are aspects of character like in the emerging adulthood period and how do they change? (2) How are the three different aspects of character connected in emerging adults? and (3) How do emerging adults see their character in the future (i.e., their character prospection)—what is important, what is ideal, and what do they plan to improve?

All three aspects of character were assessed using rating scales with anchors that are more transparent than most of the scales typically used in personality psychology, which bestows more apparent meaning on the raw scale scores. Instead of a typical agree/disagree scale or a very true of me/not true of me, all three scales used frequency scales. Moral self-attributes and character traits, because they were also being rated in other regards (that are not relevant to the current chapter), were rated on a more extensive than is typical 11-point scale. Character strivings were rated on a shorter 6-point scale. Regardless of the length of the scale, the scales' endpoints extended from "never" to "always."

The moral self-attributes scale included the six previously mentioned terms.[14] To assess character traits, a total of 20 Big Five facets were included, generally consisting of three to four adjectives: five facets for extraversion (energy, assertiveness, confidence, sociability, and optimism); four facets for agreeableness (sincerity, politeness, tolerance, and empathy), conscientiousness (dependability, caution, industry, and organization), and openness (creativity, curiosity, unconventionality and introspection); and three facets for emotional stability/neuroticism (serenity, anxiety, and unhappiness). These scales were developed from Goldberg (1990, 1992) and Saucier and Goldberg (1996), and

[13] It is also worth noting that all three aspects of character could be decontexualized or highly contextualized depending on how they are assessed.

[14] To find out more information about the moral self-attributes scale, including the response scale anchors, participant instructions, and psychometrics, please e-mail the author at enoftle@willamette.edu.

I pretested them for psychometric properties in the Eugene-Springfield community sample (see Goldberg, 2008).[15] It should be noted that for purposes of representing the Big Five broadly, not all of the 20 facet scales centrally represent character, although most do. As noted previously, trait content related to emotional stability is not usually considered part of character, although high levels of emotional stability may be considered typical of good character, if not necessary for it. In addition, other traits, specifically sociability, assertiveness, and unconventionality, may be traits that are more peripherally related to character (Lapsley & Lasky, 2001). Both moral self-attributes and the Big Five facets were rated in terms of how the person typically acts (i.e., current levels), in terms of how the person expects to act by the end of college (predicted levels), and in terms of how they would ideally want to act (ideal levels).

The 24 character strivings scales each comprised five character strength items adapted from Peterson and Seligman's (2004) VIA-IS, which were rephrased as strivings. These were rated primarily in terms of the effort devoted to the striving ("How much do you typically work on achieving this striving?"). In addition, they were also rated on the importance of the striving ("How important to you is achieving this striving?") on a six-point scale ranging from "not important" to "extremely important." Participants also reported ten personal strivings at the beginning of college, which will be briefly mentioned in this chapter to highlight possible differences between those with high versus low levels of moral self-attributes.

Participants included 237 freshman students at a western liberal arts college. The main analyses reported in this chapter are based on a long online survey they completed in their first few weeks at college as part of a longitudinal study that continued throughout their college years. Change and stability findings are based on 167 students (of those same 237) who were surveyed again at the end of their sophomore year.

CHARACTER OF EMERGING ADULTS

The first main question of the current research is about emerging adults' character at the beginning of college; what aspects of character are strongly descriptive of emerging adults and which were relatively lacking? This question was addressed in two ways. First, emerging adults' current levels of the

[15] Adjectival measures for the broad Big Five traits exist that have been thoroughly validated and are well established (e.g., Goldberg, 1992). However, very few adjectival measures of the Big Five *facets* have been developed, and therefore these haven't received as much validation attention. The two that I identified assess 18 facets (Saucier & Ostendorf, 1999) and 32 facets (Saucier, 2002), but both simply contained too many items (142 and 90, respectively) to be used in the present research, where each adjective was to be rated multiple times in the pursuit of different purposes. Thus the new 20-facet measure was developed for the current purposes, with attention to identifying distinguishable facets which represented the content of each broad trait as thoroughly as possible.

three aspects of character were identified. Second, questions about character stability and change across their first two years of college were pursued.

Table 22.2 shows how the average participants currently rated themselves at the outset of college. The findings reveal that the average participant reports frequently enacting character traits in behavior, but also falls short of always

TABLE 22.2. } **Current and Ideal Levels of Character Traits and Moral Self-Attributes**

Response scale range and anchors represented	Current levels of character traits and moral self-attributes	Ideal levels of character traits and Moral self-attributes
9–10: Between "continually" and "always"	—	Politeness (A; 9.17) Dependability (C; 9.11) Sincerity (A; 9.07)
8–9: Between "very frequently" and "continually"	—	Industry (C; 8.95) *Moral self-attributes* (8.86) Empathy (A; 8.78) Confidence (E; 8.63) Curiosity (O; 8.51) Organization (C; 8.48) Tolerance (A; 8.28) Optimism (E; 8.25) Creativity (O; 8.13) Energy (E; 8.10)
7–8: Between "quite often" and "very frequently"	Politeness (A; 7.77) Sincerity (A; 7.59) Dependability (C; 7.44) Empathy (A; 7.38) *Moral self-attributes* (7.32) Industry (C; 7.05)	Sociability (E; 7.83) Serenity (ES; 7.82) Introspection (O; 7.76) Caution (C; 7.34)
6–7: Between "often" and "quite often"	Curiosity (O; 6.89) Introspection (O; 6.68) Tolerance (A; 6.67) Caution (C; 6.63) Optimism (E; 6.50) Organization (C; 6.42) Energy (E; 6.21) Sociability (E; 6.14) Creativity (O; 6.09)	Assertiveness (E; 6.86)
5–6: Between "about half the time" and "often"	Serenity (ES; 5.98) Confidence (E; 5.89) Assertiveness (E; 5.26) Anxiety (ES; 5.17)	Unconventionality (O; 5.29)
4–5: Between "sometimes" and "about half the time"	Unconventionality (O; 4.68)	—
3–-4: Between "occasionally" and "sometimes"	Unhappiness (ES; 3.37)	—
2–3: Between "once in a while" and "occasionally"	—	Anxiety (ES; 2.48)
1–2: Between "seldom" and "occasionally"	—	Unhappiness (ES; 1.24)
0–1: Between "never" and "seldom"	—	—

Note: In parentheses is the Big Five factor linked to the character trait; (E = extraversion, A = agreeableness, C = conscientiousness, ES = emotional stability, O = openness), followed by the mean score for participants.

acting that way. Still, there's considerable variation in how the participants reported acting; most of the character traits had large standard deviations of around 1.50. In addition, there was quite a bit of variation in how the average participant acts on different traits. Character traits related to agreeableness and conscientiousness tended to be most frequently displayed in behavior; whereas traits related to extraversion and openness tended to be less frequently characteristic. The three emotional stability traits tended to be around the midpoint of the scale, and it's notable that the other traits that were suggested to be more peripherally related to character were reported as less frequent in behavior than most of the other traits (e.g., sociability).

Moral self-attributes fit in with the top group of character traits; the average participant reported acting morally somewhere between "quite often" and "very frequently." However, only a select few—a little more than 1%—reported always acting morally. As was found for the character traits, there were substantial individual differences—the standard deviation was 1.20. In fact, only 7% of the sample reported acting morally "continually" or above (< 9.0), and only about a third total reported acting morally "very frequently" or above (< 8.0). It's also worth noting, as can be seen in table 22.2, that moral self-attributes are only about as strongly endorsed on average as some commonly measured traits (e.g., dependability, empathy). Thus, social desirability biases seem not to impact moral self-attributes ratings more than traits, the latter of which have received decades of validation support using self-report methods. Such findings help alleviate concerns that everyone might see themselves as always acting morally and thus that morality self-ratings are fatally biased.

Table 22.3 demonstrates that most character strivings for the average participant were pursued somewhere between "often" and "almost always," suggesting that emerging adults' motivations were closely linked with acting with character. Interestingly, spirituality and love of learning were the character strivings rated the lowest (although still in the range of "often"). Lower than average pursuit of character strivings related to spirituality and faith may reflect that the study's population draws from one of the least religious areas of the United States (Association of Data Archives, 2010). At first blush, the lower average pursuit of love of learning is surprising given that the population under study comprised of college students, who, hopefully, should frequently pursue a love of learning. However, lower love of learning could be influenced by the high academic demands of college, as the striving refers mostly to individually motivated, independent intellectual pursuits (e.g., reading for fun), which may be more difficult to pursue in an academically rigorous setting, where so much allocated time is already devoted to intellectual pursuits demanded by coursework. In sum, it appears that for most aspects of character, people tend to see themselves as acting in good character a considerable amount of the time but not always.

TABLE 22.3 } **Current Levels of Character Strivings**

Response scale range and anchors represented	Character striving
5–6: Between "almost always" and "always"	None
4–5: Between "often" and "almost always"	Honesty (4.91) Love (4.82) Teamwork (4.81) Humor (4.74) Interpersonal intelligence (4.72) Gratitude (4.68) Kindness (4.67) Industry (4.54) Wisdom (4.53) Fairness (4.52) Appreciation of beauty (4.46) Forgiveness (4.44) Hope (4.44) Prudence (4.42) Judgment (4.41) Zest (4.39) Curiosity (4.28) Bravery (4.23) Self-control (4.22) Leadership (4.20) Humility (4.17) Originality (4.01)
3–4: Between "sometimes" and "often"	Spirituality (3.99) Love of learning (3.82)
2–3: Between "occasionally" and "sometimes"	None

Note: In parentheses are participants' mean scores for each character striving.

A second way to pursue the question of what character is like in emerging adulthood is to chart how stable these characteristics are and whether they change over the course of college. Accordingly, at the end of sophomore year, the same participants were surveyed again about all three aspects of character. The question was pursued in two ways: (1) assessing the rank-order stability of the characteristics, and (2) assessing mean-level change (Robins et al., 2001). Rank-order stability reveals how the average person changes relative to everyone else. In other words, how strongly or weakly is relative positioning maintained on a characteristic regardless of whether the group increases, decreases, or stays the same; this is assessed using correlations across the two time points. Thus, rank-order stability indexes whether the group of people tends to change in similar ways or whether people in the group tend to differ in how they change. On the other hand, mean-level change focuses on whether the group on average tends to increase, decrease, or stay the same on a certain characteristic, and compares the mean levels across the two time points. Thus, mean-level change indexes what happens to the typical person in the study over time, whether they tend to change or stay the same, and how much they change if they do change.

The rank-order stability analyses revealed that the stability of all three of the character aspects was fairly high across the first two years of college. Character trait stabilities ranged from .42 to .68, which indicates moderate to high stability; these are in the same range or a little lower than the broad Big Five traits across a comparable period (Vaidya et al., 2002). Moral self-attributes had a stability near the top of the range for the character traits, with a stability of .66, suggesting that people are indeed pretty much maintaining their relative positioning on how moral they view their own behavior, even over two years. Character strivings, even though conceptually more dynamic and changeable in nature, also had relatively high stabilities of .44 to .61. These findings demonstrate that emerging adults tend to change in relatively similar ways to one another over time. Finally, the relative positioning of the average levels of character traits and of character strivings were assessed across the two years (i.e., a profile correlation at the group level), and the respective correlations were very strong (.97 and .96), suggesting that the relative average levels between each aspects of character depicted in tables 22.2 and 22.3 were almost perfectly maintained. In other words, although emerging adults either maintain or slightly improve their character over the first two years of college, no individual character trait or striving improves dramatically as compared with the other character traits or strivings.

The mean-level change analyses suggested no radical changes in character across time, but in general a pattern of relative improvement. For the character traits, confidence and assertiveness, tolerance and empathy, organization, and curiosity and introspection all statistically significantly increased. The only significant character decline observed across the two years was for caution. Moral self-attributes did not significantly increase, although there was a positive trend upward. Six of the 24 character strivings significantly increased: curiosity, fairness, industry, judgment, leadership, and originality; no character strivings decreased. Thus, the typical person in the study tended not to improve that much on the three aspects of character overall, but did improve for some of the attributes. However, two years is a relatively short period of time; across a longer period of time more pervasive mean-level changes in traits may perhaps be observed.

One possible concern with the current approach so far is that even though the ratings are on a transparent rating scale and can be interpreted directly, the desirability of the attributes under study might inflate how often they are reported, leading to biased estimates. However, it's worth pointing out that there are considerable individual differences in how much people report behaving in good character, show effort in striving toward good character, and view themselves to be acting morally. Sizable individual differences suggest meaningful differences among people, especially since they are rather stable over time. In addition, very few participants rated themselves at the top of the scales, which would have reflected desirability bias, and improvements are

indeed still possible to observe over time. However, it's still certainly possible that desirability generally inflated people's ratings, although, to be sure, this is the case with other self-report scales. Some of the further research questions address concerns about desirability, for example, by assessing ideal levels, a question pursued later, in the section, "Character Prospection: Ideals, Improvement, and Importance."

In sum, the view of character in emerging adulthood is quite positive, as might be expected from the personality development literature on the Big Five traits. Not only do emerging adults' current behaviors and strivings reflect a great deal of focus on good character, but such tendencies seem to be maintained or even improved across the first two years of college. Such findings would seem to argue against young people as excessively narcissistic (Arnett, 2010), and would suggest emerging adulthood is a time of focus on both character and character development.

CONNECTIONS BETWEEN ASPECTS OF CHARACTER

The second main question of the current research is to see how strongly or loosely the different aspects of character connect to one another. Does seeing oneself as acting morally also mean that one displays more character traits and pursues more character strivings?[16] This question was pursued by examining the basic associations between the three sets of character constructs.

In general, moral self-attributes were moderately connected to character traits. Interestingly, the traits included that were conceptually more peripheral to character also were more weakly related to the moral self-attributes. For example, unconventionality and anxiety were simply uncorrelated with moral self-attributes (r's of .02 and .06, respectively). Unhappiness was modestly negatively correlated ($r = -.19$), and serenity, sociability, and assertiveness were modestly positively correlated (r's between .17 and .25; p's $< .05$). However, the rest of the traits, which definitionally are more central to character, were moderately to strongly correlated with moral self-attributes, ranging from .32 to .71. These character traits are clearly substantially associated with seeing oneself as a moral person. In particular, the character traits related to Agreeableness and Conscientiousness tended to be most strongly associated to moral self-attributes. The same trend found for character traits was also

[16] The current research is based on self-reports, so it is possible that seeing oneself as acting morally predicts only *seeing oneself* as displaying more character traits and pursuing more character strivings. However, at least for character traits, a century of research suggests that self-reports of traits are valid, in that they are associated closely with informant reports of traits (John & Robins, 1993b; McCrae, 1982; Watson, Hubbard, & Wiese, 2000) and life outcomes (Ozer & Benet-Martinez, 2006), among other criteria. Some similar validation has been found for character strengths (Buchor, Proyer, & Ruch, 2013; Ruch et al., 2010), if not yet for the new formulation of character strivings.

found between moral self-attributes and character strivings. All 24 character strivings were statistically significantly positively associated with moral self-attributes, ranging from modest correlations with originality ($r = .25$) to strong associations with prudence ($r = .58$).

Although they're not a focus of this chapter, the current research also asked participants to name ten personal strivings. To further demonstrate the connection between moral self-attributes and strivings, it might be illustrative to provide a taste of what was observed from the personal strivings. Table 22.4 shows the set of personal strivings for two participants who have among the highest levels of moral self-attributes and two with among the lowest levels. The difference between the two sets of personal strivings is fairly striking. Those high in moral self-attributes appear less focused on appearance and performance, seem to pursue more realistic, prosocial strivings, yet also maintain a sense of autonomy, as would be predicted by Hogan. On the other hand, those low in moral self-attributes appear to be generally preoccupied with how they seem to others, offer negativistic and self-defeating personal strivings, and seem either overly dependent on others or conversely highly manipulative, but hardly autonomous.

As might be expected from past research connecting personality traits and character strengths (Noftle et al., 2011), character traits were quite strongly related to character strivings, and vice versa. The strongest correlations between each of the 24 character strivings and the closest associated character trait ranged from .45 to .69, suggesting a good amount of overlap but not full redundancy. As with Noftle et al., it also was demonstrated that a sizable number of the character traits are connected to strivings. In fact, 12 different character traits (of the 20 traits in total) were represented in this empirical mapping of each of the 24 character strivings to traits: four extraversion traits, three conscientiousness traits, three agreeableness traits, and two openness traits. In addition, four other traits that were not among this 12-trait list, confidence, politeness, organization, and introspection, demonstrated correlations of at least .40 with at least one of the character strivings. The remaining four traits of the 20, the three traits related to emotional stability, as well as unconventionality, were more weakly related to the character strivings, which is unsurprising due to their definitionally more peripheral status within character.

In sum, the results on associations between aspects of character suggest that all three tend to be moderately to strongly connected in emerging adults, but none of them to the degree that they appear to completely overlap with one another. Thus, they appear to be assessing different but important features of character. These patterns of sizable and meaningful correlations, as well as the telling differences in the personal strivings of those with high versus low moral self-attributes, argue that this new construct may have been overlooked in previous research but that it has the future potential to contribute to the

TABLE 22.4 } **Personal Strivings of Sample Individuals with High and Low Levels of Moral Self-Attributes**

Sample high moral self-attributes raters (from top 5%)		**Sample low moral self-attributes raters (from bottom 5%)**	
Participant 128A Score = 10.0	Participant 003A Score = 9.67	Participant 019C Score = 4.17	Participant 039A Score = 3.83
Be understanding of others	Be a good listener	Be perfect or seem perfect to others	Be physically attractive
Do my best work in all of my classes	Be active in my community	Achieve my best or the best in someone else's eyes	Be seen as charas-matic [*sic*]
View experiences in a positive light even when things go wrong	Be a beneficial asset to my community.	Lose weight	Be funny
Learn something new every day	Find people who have different opin-ions than mine	Achieve my goals	Succeed in my classes
Make time for friends and family as well as academics	Encourage others in whatever way possible	Learn as much as possible	Be seen as smart
Make time every day for myself, for relax-ation and exercise	Be as kind as I can to everyone	Manipulate the situation	Get ahead
Tolerate differ-ent opinions while remaining true to my own beliefs	Spend time reflecting on my actions and my thoughts	Control my life	Develop unique skills
Be open to new expe-riences but not take unnecessary risks	Read books, articles, and other snippets that interest me	Control others' lives	Make my friends happy
Look at ideas from multiple or opposing points of view	Display the love God has for me through my words and actions	Defy others' expectations	Continue on even when i should take a break
Be conscious of my actions and their effects on others	Be who I believe myself to be, not what others believe I should be	Surprise people	Shoulder all of the stress

study of character not only as an abstracted feature of a moral exemplar but also as a self-attribute.

CHARACTER PROSPECTION: IDEALS, IMPROVEMENT, AND IMPORTANCE

The third main question of the current research has to do with character prospection, or how emerging adults view the aspects of character in the future in terms of aspirations and motivation. This question was addressed somewhat differently for the three aspects of character. Character traits and

moral self-attributes both are connected with particular levels of behavior, which easily connect to aspirations: how a person expects to change and also how he or she would ideally act. On the other hand, character strivings have to do with recurring motivations, and thus may differ mainly in how much focused attention is paid to enacting them (as already discussed) but also in how much importance is placed on them. It is less certain what an ideal level of a striving might reflect, and people may not think much about changing the *levels* of their strivings, but simply about what strivings they have. Because of these differences, the question of prospection is addressed in three ways. One, emerging adults were asked about their ideal levels for character traits and moral self-attributes. Two, emerging adults were asked about how much they expected to change on these characteristics across a tangible time period in the near future: the college years. Three, emerging adults were asked about the relative importance of achieving the 24 character strivings, providing a glimpse into which strivings they might be most motivated to pursue in the near future.

One way in which character prospection was addressed was by asking participants about their ideal levels of character traits and moral self-attributes, which helps to determine whether participants see their current attributes as close to ideal or in need of some improvement. In addition to depicting the current levels of character traits and moral self-attributes, table 22.2 also shows emerging adults' ideal levels. At least four things are notable. First, except for anxiety and unhappiness, which are understandably ideally lower than the current levels, all ideal levels for character traits are higher than current levels. Indeed, some of the ideal levels are considerably higher, and most amount to a scale point or two higher. This may allay major concerns about social desirability; people are not simply describing their current behavior as how they would ideally see it. Second, although ideal levels are quite high, they're interestingly also not all at the top of the scale. This suggests some interesting implications. Perhaps most emerging adults believe that acting in good character does not demand acting polite, dependable, or confident all the time but, instead, *almost* all the time. Maybe they envision relatively rare situations in which one would be morally obligated not to be polite (e.g., defending a close other from harm), or when it would be better not to be dependable (e.g., following orders that are ethically questionable) or not to be confident (e.g., when others know more than you do). Alternatively, perhaps most emerging adults do not believe it is ideal to *always* act in good character, just most of the time. The truth is likely some combination of these and maybe other possibilities. Although ideal levels for moral self-attributes themselves on average were quite high at 8.86, higher than most of the character traits, only 13% of the sample placed ideal levels of moral self-attributes at the very top of the scale, and only a slight majority (51%) put their ideal levels at 9 or above (i.e., acting that way continually).

A third notable finding from the examination of ideal levels was that there was a trend for the traits that have a higher frequency of behavior to also have higher ideal levels. In other words, people reported that they typically acted polite, sincere, and dependable very frequently in their current behavior (i.e., more often than some of the other character traits), and also placed their ideal levels of these characteristics out on top as well (i.e., higher than ideal levels of other character traits). Fourth, ideal-current discrepancies, representing the direction and strength of desired changes, revealed something else altogether. When the highest *discrepancies* between current and ideal behavior were examined, interesting results emerged that didn't fit with the trend of similar rank orders between current and ideal levels. Neuroticism-related traits, most peripheral to character, were close to the top of the list for ideal-current discrepancies, but had been lower in current levels. Ideally, people wanted to be much less anxious and unhappy. In addition, it wasn't the most central character traits with the highest current levels that were most desired to change; instead, it was confidence, organization, creativity, industry, and energy. Moral self-attributes were at the middle of the pack. Part of this reflects that those with higher current levels have less potential distance to travel on average until they get to an ideal, but as seen earlier, even ideal levels do not completely max out the possible scale. However, this is not to say that more central character traits are not desired to change much. Only caution and unconventionality had discrepancies of less than a scale point in size, in part due to the fact that the former included a sizable minority who wanted to ideally be *lower* on caution, and the latter included a sizable minority who didn't want to change their current levels. However, in Cohen's *d* units, which gage the magnitude of an effect size, most discrepancies between current and ideal levels range from a moderate to well over a large difference in magnitude (Cohen, 1992). This all suggests that although people's current behavior reflects a great deal of positive character, people wish it reflected considerably more.

A second way character prospection was addressed to determine future expectations for change was by asking emerging adults about how they predicted their character traits and moral self-attributes to be at the end of college. Just as taking discrepancies between ideal and current levels identifies how much people would ideally desire to change, taking discrepancies between predicted levels and current levels identifies people's expectations for change within a certain time frame. Predictably, each of these expectations was smaller than people's ideal desires on the matched trait. However, which character traits did participants expect to change the most on? The largest expected positive changes were for the following traits, in descending order: confidence, organization, industry, sociability, and assertiveness. In particular, these are the somewhat self-focused, agentic traits that are most relevant to developmental tasks of emerging adults: forging social connections, establishing independence, and striving toward a career (Arnett, 2000). These traits cohere

strongly with Hogan's (1973) identification of autonomy as the focal aspect of moral development during this period. On the other hand, traits that were more central to character, and most strongly related to moral self-attributes, were further down the list—as was moral self-attributes itself. Such character traits were still expected to change across college, and to change at non-trivial levels, but were not prioritized quite so strongly in expected change.

A third way in which character prospection was addressed was by asking participants about the relative importance of character strivings. Although participants were not asked about what they thought their character strivings would be like in the future, the aim was to reveal through ratings of importance which character strivings might be most highly prioritized in coming years. On average, all the character strengths were held to be important, as ratings ranged approximately from 4.5 to 5.5 on the six-point scale, thus between "moderately important" and "extremely important." Interestingly, a different pattern of content emerged from these analyses than from the previous ones on character prospection, with a focus much more on character strivings having to do with communal than on agentic, self-directed content. The character strivings that were held to be most important were, in descending order: honesty, gratitude, love, hope, interpersonal intelligence, kindness, humor, and fairness.

In sum, findings from the third question about character prospection suggest that although emerging adulthood is a time when a lot of attention is paid to having good character, emerging adults also both desire and expect improvements in character. In particular, they ideally desire quite a bit of improvement across character traits and moral self-attributes. Their outright expectations for change across the college years are more modest, but also seem to highlight the traits that are most relevant to age-graded developmental tasks rather than the traits most central to character. However, in terms of strivings, emerging adults seem to prioritize those character qualities that are highly prosocial and communal in nature. These findings demonstrate that a great deal is learned from character prospection that cannot be observed from studying current character levels alone.

Future Directions

Clearly, there is a lot more to discover about these three aspects of character in emerging adulthood. First, although the results suggest that emerging adults as a group report having good character and seem to be developing in positive ways, future research should pursue at least three potential limitations: a focus on only some aspects of character, limited generalizability, and reliance on self-reports. In terms of the first issue, although the current research included three different aspects of character, more aspects and approaches exist. For example, important

aspects of character are captured not only within the domains of traits and characteristic adaptations (e.g., Lapsley & Hill, 2009), but also within a third type of construct in personality, that of narrative identity (McAdams, 2009). In addition, it will be important to examine character from a state approach (e.g., using experience sampling methods; Fleeson & Noftle, 2011) alongside a dispositional approach (Fleeson, 2012; Hooker, 2002). In terms of the second issue, the current sample of college students at a premier liberal arts college is likely not representative of emerging adults in general. In terms of the third issue, although self-reports of some aspects of character have been well validated, the same is not true of the new formulations of character strivings and moral self-attributes. Moreover, even though informant reports and self-reports tend to agree, the two perspectives are important sources of information that tend not to be redundant (Vazire, 2010), and both will be useful in studying character.

Second, although the three aspects of character have been found to generally be related, more needs to be known about how they're related: how do they emerge developmentally, and how do they work together in terms of social-cognitive processes linking them together functionally, connecting the descriptive aspects of the characteristics with their explanatory underpinnings, as is the goal in whole trait theory (Fleeson, 2012). In particular, as moral self-attributes are a newly studied construct, more needs to be understood about them—for example, their links to moral identity—before their usefulness in understanding character is fully demonstrated.

Third, character prospection suggests that emerging adults are both far from their ideal levels of character and also hold great expectations for change over a relatively short period of time, which, on the one hand, might be unrealistic but on the another hand, might fuel actual change. More needs to be known about the functions and outcomes of character prospection (Noftle, in preparation).

Conclusion

Although more work needs to be done, the current results suggest that emerging adulthood is a fruitful time during which to pursue the study of character. Emerging adults not only typically report exhibiting good character frequently in their behavior, they also report chronically striving toward a variety of virtuous qualities, which they deem to be quite important. They view themselves as acting morally frequently but not always. Although character seems to be rather stable in the first half of college, it also shows a few targeted improvements. On the other hand, emerging adults also demonstrate some humility indicative of good character: they reveal that they ideally would desire even higher levels of character traits and moral self-attributes, and have targeted expectations for change across college in positive ways that seem to most

closely fit the developmental tasks of emerging adulthood. These three diverse aspects of character come together coherently in early emerging adults and promise some exciting future developments in further study.[17]

References

Allport, G. W. (1931). What is a trait of personality? *Journal of Abnormal and Social Psychology*, *25*, 368–372.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York, NY: Henry Holt.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*, 1–211.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC.

Armon, C., & Dawson, T. L. (1997). Developmental trajectories in moral reasoning across the life span. *Journal of Moral Education*, *26*, 433–453.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*, 469–480.

Arnett, J. J. (2010). Oh, grow up! Generational grumbling and the new life stage of emerging adulthood: Commentary on Trzesniewski & Donnellan (2010). *Perspectives in Psychological Science*, *5*, 89–92.

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*, 1531–1544.

Ashton, M. C., & Lee, K. (2005). Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, *73*, 1321–1353.

Association of Data Archives. *All denominations: Rates of adherence per 1000 population*. (2010). Retrieved from http://www.thearda.com/mapsReports/maps/Ardamap.asp?GRP=1&map1=3

[17] Parts of this project were made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. I would like to acknowledge several key parties in the current research. First, I would like to express heartfelt gratitude to the participants in the current research for their dedication to an intensive, and at times perhaps exhausting, multipart, multiyear, longitudinal study. Second, I would like to offer my deepest thanks to the current and former members of the Willamette University Personality ("WUP") Lab for their intelligent, invaluable help during the course of the longitudinal study: Erika Rauch, Rachel Woods, Ariel Wilson, Kellie Menghini, Kayla Kosaki, Brian Gnerre, Megan Nanry, Charleen Gust, Nina Wiester, Stephanie Chan, Cody Christensen, and Madison Montemayor. I couldn't have conducted the study without your help. Third, I am indebted to and very appreciative of Kendra Mingo, Michael Moon, Jamie Hart, Laura Jacobs Anderson, David Douglass, Lisa Holliday, Ross Stout, and the Willamette psychology department and Dean's Office for their enthusiastic and enduring support at the institutional level. Fourth, I would like to thank and acknowledge Christian Miller, Will Fleeson, and Mike Furr for their very thoughtful comments on an earlier version of this chapter, as well as thank them and the other attendees at the Character Project Workshops in 2012 and 2013 for their highly useful feedback as the research was progressing. Finally, I would like to thank and commend the John Templeton Foundation for their continuing funding of scientific work on character and character development, as well as the Character Project itself.

Buchor, C., Proyer, R. T., & Ruch, W. (2013). Self- and peer-rated character strengths: How do they relate to satisfaction with life and orientations to happiness? *Journal of Positive Psychology*, *8*, 116–127.

Cawley, M. J., Martin, J. E., & Johnson, J. A. (2000). A virtues approach to personality. *Personality and Individual Differences*, *28*, 997–1013.

Cantor, N. (1990). From thought to behavior: "Having" and "doing" in the study of personality and cognition. *American Psychologist*, *45*, 735–750.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Davis, M. H., & Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *Journal of Research in Personality*, *25*, 70–87.

De Raad, B., & Van Oudenhoven, J. P. (2011). A psycholexical study of virtues in the Dutch language, and relations between virtues and personality. *European Journal of Personality*, *25*, 43–52.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, *124*, 197–229.

Emmons, R. A. (1986). Personal strivings: An approach to personality and subjective well-being. *Journal of Personality and Social Psychology*, *51*, 1058–1068.

Emmons, R. A. (2003). Personal goals, life meaning, and virtue: Wellsprings of a positive life. In C. L. M. Keyes & J. Haidt (Eds.), *Flourishing: Positive psychology and the life well-lived* (pp. 105–128). Washington, DC: American Psychological Association.

Fleeson, W. (2012). Perspectives on the person: Rapid growth and opportunities for integration. In K. Deaux & M. Snyder (Eds.), *The Oxford handbook of social and personality psychology* (pp. 33–63). New York, NY: Oxford University Press.

Fleeson, W., & Heckhausen, J. (1997). More or less "me" in past, present, and future: Perceived lifetime personality during adulthood. *Psychology and Aging*, *12*, 125–136.

Fleeson, W., & Noftle, E. (2008). The end of the person-situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*, 1667–1684.

Fleeson, W., & Noftle, E. E. (2011). Personality research. In M. R. Mehl, & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 525–538). New York, NY: Guilford.

Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, *2*, 31–39.

Goldberg, L. R. (1990). An alternative "description of personality": The big five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216–1229.

Goldberg, L. R. (1992). The development of markers for the big five factor structure. *Psychological Assessment*, *4*, 26–42.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34.

Goldberg, L. R. (2008). The Eugene-Springfield community sample: Information available from the research participants. *Oregon Research Institute Technical Report 48*(1), 1–19. http://projects.ori.org/lrg/PDFs_papers/ORI_TechnicalReport_ESCS_Mar08.pdf

Hartshorne, H., May, M. A., Maller, J. B., & Shuttleworth, F. K. (1928). *Studies in the nature of character*. New York, NY: MacMillan.

Haslam, N., Bastian, B., Fox, C., & Whelan, J. (2007). Beliefs about personality change and continuity. *Personality and Individual Differences*, *42*, 1621–1631.

Haase, C. M., Heckhausen, J., & Wrosch, C. (2013). Developmental regulation across the life span: Toward a new synthesis. *Developmental Psychology*, *49*, 964–972.

Hennecke, M., Bleidorn, W., Denissen, J. J. A., & Wood, D. (2014). A three-part framework for self-regulated personality development across adulthood. *European Journal of Personality*, *28*, 289–299.

Hill, P. L., & Roberts, B. W. (2010). Propositions for the study of moral personality development. *Current Directions in Psychological Science*, *19*, 380–383.

Hogan, R. (1973). Moral conduct and moral character: A psychological perspective. *Psychological Bulletin*, *79*, 217–232.

Hogan, R., Johnson, J. A., & Emler, N. P. (1978). A socioanalytic theory of moral development. In W. Damon (Ed.), *New directions for child development: Vol. 2. Moral development* (pp. 1–18). San Francisco, CA: Jossey-Bass.

Hooker, K. (2002). New directions for research in personality and aging: A comprehensive model for linking levels, structures, and processes. *Journal of Research in Personality*, *36*, 318–334.

Hooker, K., & McAdams, D. P. (2003). Personality reconsidered: A new agenda for aging research. *Journals of Gerontology: Series B. Psychological Sciences and Social Sciences*, *58*, P296–P304.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). New York, NY: Guilford.

John, O. P., & Robins, R. W. (1993a). Gordon Allport: Father and critic of the five-factor model. In K. H. Craik, R. T. Hogan, & R. N. Wolfe (Eds.), *Fifty years of personality research* (pp. 215–236). New York, NY: Plenum.

John, O. P., & Robins, R. W. (1993b). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*, 521–551.

Kohlberg, L. (1981). *The meaning and measurement of moral development*. Worcester, MA: Clark University Press.

Krueger, J. & Heckhausen, J. (1993). Personality development across the adult life span: Subjective conceptions vs. cross-sectional contrasts. *Journal of Gerontology: Psychological Sciences*, *48*, P100–P108.

Lapsley, D. K. (1996). *Moral psychology*. Boulder, CO: Westview Press.

Lapsley, D. K., & Hill, P. T. (2009). The development of the moral personality. In D. Narvaez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 185–213). New York, NY: Cambridge University Press.

Lapsley, D. K., & Lasky, B. (2001). Prototypic moral character. *Identity: An International Journal of Theory and Research*, *1*, 345–363.

Leary, M., Tambor, E., Terdal, S., & Downs, D. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology*, *68*, 518–530.

MacDonald, C., Bore, M. R., & Munro, D. (2008). Values in action scale and the big 5: An empirical indication of structure. *Journal of Research in Personality*, *42*, 787–799.

McAdams, D. P. (2009). The moral personality. In D. Narvaez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 11–29). New York: Cambridge University Press.

McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, *43*, 293–303.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175–215.

McCullough, M. E., & Snyder, C. R. (2000). Classical sources of human strength: Revisiting an old home and building a new one. *Journal of Social and Clinical Psychology*, *19*, 1–10.

McGrath, R. E. (2014). Scale- and item-level factor analyses of the VIA Inventory of Strengths. *Assessment*, *21*, 4–14.

Miller, C. B. (2014). *Character and moral psychology.* New York, NY: Oxford University Press.

Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, *61*, 622–631.

Nicholson, I. A. M. (1998). Gordon Allport, character, and the "culture of personality," 1897–1937. *History of Psychology*, *1*, 52–68.

Noftle, E. E. (2010). The future expectations and desires of emerging adults in self-motivated, intentional personality change. Talk presented at the 15th meeting of the European Conference on Personality. July. Brno, Czech Republic.

Noftle, E. E. (in preparation). *From amateur to auteur: Future expectations and desires for personality change in emerging adulthood*. Manuscript.

Noftle, E. E., Schnitker, S. A., & Robins, R. W. (2011). Character and personality: Connections between positive psychology and personality psychology. In K. Sheldon, T. Kashdan, & M. Steger, (Eds.), *Designing the future of positive psychology: Taking stock and moving forward* (pp. 207–227). New York, NY: Oxford University Press.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*, 574–583.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421.

Paulhus, D. L., & Williams, K. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556–568.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the big five? Plenty! *Journal of Personality*, *68*, 821–835.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook of classification*. Washington, DC, and New York: American Psychological Association and Oxford University Press.

Plant, W. T., & Telford, C. W. (1966). Changes in personality for groups completing different amounts of college over two years. *Genetic Psychology Monographs*, *74*, 3–36.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A metaanalysis of longitudinal studies. *Psychological Bulletin*, *132*, 1–25.

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, *69*, 617–640.

Robins, R. W., Tracy, J. L., Trzesniewski, K. H., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality*, *35*, 463–482.

Ruch, W., Proyer, R. T., Harzer, C., Park, N., Peterson, C., & Seligman, M. E. P. (2010). Adaptation and validation of the German Version of the Values in Action Inventory of Strengths (VIA-IS) and the development of a peer-rating form. *Journal of Individual Differences*, *31*, 138–149. doi:10.1027/1614-0001/a000022

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the big five. *Journal of Research in Personality*, *36*, 1–31.

Saucier, G., & Goldberg, L. R. (1996). Evidence for the big five in analyses of familiar English personality adjectives. *European Journal of Personality*, *10*, 61–77.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the big five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, *76*, 613–627.

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet-level analysis. *Personality and Social Psychology Bulletin*, *30*, 1062–1075.

Schnitker, S. A., & Emmons, R. A. (2007). Patience as a virtue: Religious and psychological perspectives. *Research in the Social Scientific Study of Religion*, *18*, 177–207.

Seligman, M. E. P., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, *8*, 119–141.

Steger, M. F., Hicks, B. M., Kashdan, T. B., Krueger, R. F., & Bouchard, T. J., Jr. (2007). Genetic and environmental influences on the positive traits of the Values in Action classification, and biometric covariance with normal personality. *Journal of Research in Personality*, *41*, 524–539.

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2008). Do today's young people really think they are so extraordinary? An examination of secular changes in narcissism and self-enhancement. *Psychological Science*, *19*, 181–188.

Twenge, J. M., & Campbell, W. K. (2009). *The narcissism epidemic: Living in the age of entitlement*. New York, NY: Free Press.

Vaidya, J. G., Gray, E. K., Haig, J., Mroczek, D. K., & Watson, D. (2008). Differential stability and individual growth trajectories of big five and affective traits during young adulthood. *Journal of Personality*, *76*, 267–304.

Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology*, *83*, 1469–1484.

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*, 281–300.

Walker, L. J., & Frimer, J. A. (2009). In D. Narvaez & D. K. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 232–255). New York, NY: Cambridge University Press.

Walker, L. J., & Pitts, R. C. (1998). Naturalistic conceptions of moral maturity. *Developmental Psychology*, *34*, 403–419.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, *78*, 546–558.

Wood, D. & Roberts, B. W. (2006). The effect of age and role information on expectations for big five personality traits. *Personality and Social Psychology Bulletin*, *32*, 1482–1496.

23 }

Etiquette and Exemplarity in Judaism

Tzvi Novick

In this chapter I advance an argument for the structural significance of etiquette in the normative space of traditional rabbinic Judaism (in the continuation, simply "Judaism").[1] The argument divides into four sections. Insofar as Jewish normativity centers on law, it appears to have trouble making claims to shaping character, or put differently, it allows exemplars to play little or no apparent role (see section 1, "The Problem"). But traditional Jewish sources do make frequent use of exemplars in the domain of etiquette (section 2, "Exemplarity and Etiquette in Judaism"). Etiquette combines, distinctively, two normative features not usually found together—namely, an intense interest in rule following and a prominent role for exemplarity (section 3, "Etiquette as a Normative Form"). In virtue of this distinctive coupling, etiquette plays an important role in enabling Jewish normativity both to remain law centered and simultaneously to purport to shape character (section 4, "The Solution").

My argument aims to contribute to the theology of character in Judaism, and in this regard two clarifications are in order. First, my analysis is descriptive, not constructive. That is, my object of study is Jewish theology as it has been, and as it is, not as it ought to be. Second, because the disciplinary category of theology originates in Christianity, to speak of Jewish theology is necessarily to deploy the category in a different and in some respects unfamiliar way. Perhaps the most dramatic difference, and certainly the most relevant for this chapter, concerns the aforementioned centrality of law. Jewish theology in most of its traditional forms assigns to law, or the law, or the Torah, an

[1] I leave the object of interest, "traditional rabbinic Judaism," intentionally vague. On any definition, traditional rabbinic Judaism is not monolithic, and the claims that I make for it are truer for some periods and places than for others. I am thinking chiefly of some varieties of Orthodox Judaism in the modern period, and the ways in which they read the classical rabbinic corpus of late antiquity, but I think that the argument holds, *mutatis mutandis*, for significant strains in the classical corpus itself, and in the medieval period.

importance not entirely unlike that which Christ assumes in Christian theology. This fact is the starting point of my argument.

1. The Problem

Ze'ev Maghen, a Jewish scholar who specializes in Islam, has argued that Judaism and Islam, unlike Christianity and other religions, self-consciously refrain from shaping character. The overarching normative framework for Judaism and Islam is not ethics but law: *halakhah* in Judaism, *shari'ah* in Islam. Unlike ethics (at least on some accounts), law works "to police people's actions externally, not . . . to reform or rewire people internally. This 'non-interference policy' allows the human being's internal drives to remain fierce; it reins them in for the sake of society, but leaves them intact for the sake of the individual."[2] In Maghen's apology for Judaism and its cousin religion, Jewish *halakhah* and Islamic *shari'ah* affirm and preserve human wildness by governing it from the outside, by fencing it in rather than breaking it.

It is easy to find counterexamples to Maghen's thesis. Judaism—and in the continuation I will address Judaism alone—boasts a rich ethical literature that extends from the Hebrew Bible to the classical rabbinic canon to modern Jewish philosophy, and directly engages with the categories of character and virtue.[3] But because law has indeed defined and in traditional circles continues to define Jewish normative discourse, to rest one's case on the relatively peripheral corpus of ethical literature is already to concede Maghen's point. A richer response to Maghen—and I do not want altogether to refute his thesis but to complicate it—comes from closer consideration of the ways in which law-like normativity interacts with other forms of normativity in traditional Judaism.

We can restate Maghen's claim in terms of exemplarity: The normative space of Judaism is dominated by law, and the domain of law allows little room for exemplarity. Exemplarity involves the thing exemplified and the exemplar. It functions robustly when the exemplar allows her audience to access, in one sense or another, the thing exemplified. The thing exemplified may be abstract, and the exemplar helps to concretize it. Or it may be vague or ambiguous, and the exemplar helps to specify it. Alternatively, the thing exemplified may be daunting to achieve, and the exemplar demonstrates that achievement is possible. Amy Olberding, in her recent exemplarist approach to the *Analects*, builds on the work of Linda Zagzebski to identify an even stronger role for the exemplar at the foundation of Confucianism. For Zagzebski and Olberding, the

[2] Maghen 2011: 236. By "law" Maghen has in mind first and foremost ritual law. Jewish civil law and criminal law, like their modern statist counterparts, more or less transparently incorporate and attempt to inculcate a particular vision of virtue and vice. On criminal law and virtue, see Duff 2006.

[3] For an excellent overview, see Mittleman 2012.

exemplar is conceptually prior to the thing exemplified. Our understanding of courage, for example, begins with, and always remains tethered to, an original act of pointing out courageous exemplars. Our conceptual understanding of courage represents only a provisional interpretation of these exemplars. This interpretation never exhausts the exemplars, and is always subject to revision as a result of further study of the exemplars.[4]

Exemplars do not, in general, have any such place in a legal framework. An exemplar cannot, of course, be conceptually prior to the laws, in the sense that Zagzebski and Olberding argue for with respect to virtues. Laws acquire their normative force from authorized legislative bodies, not from exemplars.[5] Because laws ordinarily take the form of rules rather than principles, there can be little pedagogical payoff from an exemplar. We do not need an exemplar to clarify what respecting our neighbor's property line means in the way that we might look to exemplars to illustrate what it means to respect our parents. Finally, laws generally make relatively minimal demands upon us. It does not take a hero to be law-abiding in more or less the fullest sense, nor does law need a hero to inspire law-abidingness.

One way of making room for exemplarity in Jewish normative space is to conceive of the law as a regimen for training in virtue. This move has a long pedigree in pre-rabbinic Judaism. Indeed, from the earliest point at which Judaism became, to an important degree, a religion of law, the law was understood, *pace* Maghen, as an expression of the highest virtue, and the study and practice thereof a path toward such virtue. It is this understanding that informs such attempts to unify the law ("Torah") and wisdom as one finds in Deuteronomy (seventh century BCE), in the book of Ben Sira (third century BCE), and in Philo (first century CE) and his philosophical successors.[6] But mainstream rabbinic Judaism commits itself fairly strongly to resisting such attempts. The reason emerges clearly from the following passage from Philo on the commandment of circumcision.

> There are some who, regarding laws in their literal sense in the light of symbols of matters belonging to the intellect, are overpunctilious about the latter, while treating the former with easy-going neglect. Such men I for my part should blame for handling the matter in too easy and off-hand a manner: they ought to have given careful attention to both aims, to a more full and exact investigation of what is not seen, and in what is seen to be stewards without reproach. As it is, as though they were living alone by themselves in a wilderness, or as though they had

[4] See Olberding 2012.

[5] I do not mean to claim that the imprimatur of an authorized legislative body is a sufficient condition, only that it is a necessary one.

[6] On Deuteronomy, see Weinfeld 1972. On Ben Sira and Philo, see Collins 1997. On Philo's most important philosophical successor, Maimonides, see Halbertal 2014: 148–58.

> become disembodied souls, and knew neither city nor village nor household nor any company of human beings at all, overlooking all that the mass of men regard, they explore reality in its naked absoluteness. These men are taught by the sacred word to have thought for good repute, and to let go nothing that is part of the customs fixed by divinely empowered men greater than those of our time. . . . It is true that receiving circumcision does indeed portray the excision of pleasure and all passions, and the putting away of the impious conceit . . . but let us not on this account repeal the law laid down for circumcising. Why, we shall be ignoring the sanctity of the Temple and a thousand other things, if we are going to pay heed to nothing except what is shown us by the inner meaning of things. Nay, we should look on all these outward observances as resembling the body, and their inner meanings as resembling the soul. It follows that, exactly as we have to take thought for the body, because it is the abode of the soul, so we must pay heed to the letter of the laws.[7]

Philo engages in this passage with fellow Alexandrian Jews who, like Philo himself, read the commandment of circumcision allegorically, as an injunction to excise "pleasure and all passions." He rebukes them for going so far as to entirely disregard the literal sense, that is, to forgo circumcision in the flesh. Philo counters this antinomian stance, first, with the contention that because even philosophers live in society, they too must heed its conventions. Second, as one must take care of the body not for its own sake but because it is the abode of the soul, so the letter of the law, insofar as it is a vehicle for the spirit, carries force. But these pragmatic and analogical considerations represent only superficial rebuttals to the elegant simplicity of Philo's opponents' antinomian position.[8]

Traditional Judaism therefore looks suspiciously on attempts such as Philo's to recast law as training in virtue. But it does not give up on the connection between law and ethics. It continues to assert a connection, but is ideologically committed to not specifying it. It is not so much, then, that Maghen is right to deny Judaism's interest in character formation. Rather, he puts the spotlight on an inherent tension in Judaism, which wants both to construct the law as law, to be followed because it is God's command, but also to have the law in fact play a role in character formation. It wants the law to be heteronomous—"to police people's actions externally," in Maghen's words—but also embodied.

---

[7] Colson and Whitaker 1939: 89–93. On Philo's targets in this passage, see Niehoff 2011: 175 and *passim*, and for a similar critique, possibly by an Alexandrian Jewish intellectual who lived roughly one hundred years after Philo, see Niehoff 2013: 171.

[8] On Philo's successors see Heinemann 2008. Perhaps the best-known passage expressing resistance to the identification of a law's ethical basis is bBerakot 32b, where one who praises God's mercy in commanding that the mother bird be shooed away before one collects her eggs (Deut. 22:6) is criticized because "he makes the attributes of the Holy Blessed One into mercies when they are really decrees." For this translation, with discussion, see ibid., 23.

This tension can be reframed again in terms of the figure of the rabbi. The basic qualification for the rabbinate in traditional Judaism is expertise in ritual law. An ordained rabbi must be able to field questions about whether a particular action represents work of a sort forbidden on the Sabbath, whether a pan used with such-and-such ingredients under such-and-such conditions has run afoul of the laws of *kashrut*, and so on. Rabbis with additional expertise in civil law serve on courts, adjudicating cases of divorce and applying Jewish civil law to financial disputes. In the realm of character, his role demands, in itself, no more than the sort of integrity one looks for in a judge.[9] At the same time, the rabbi, as a professional student of the law, is supposed to embody virtue. The relationship between his acknowledged legal expertise and virtue is, in general, left unspecified and obscure.

2. Exemplarity and Etiquette in Judaism

I want to begin to mount an argument about one way in which Judaism relieves this tension by highlighting an area in which canonical Jewish sources from late antiquity forward make regular and profound use of exemplarity, namely, etiquette. Rabbinic literature puts forward rabbis as authorities on matters of etiquette, and it is often their actions rather than or in addition to their reasoned views that justify the rule. One example (yShabbat 6:2 [8a–b]) concerns the proper procedure for binding one's sandals. Shimon bar Ba, serving Rabbi Yohanan, approaches him with the sandal for the right foot. He does so in obedience to a treatise on etiquette (*derek eres*) that specifies, in accordance with the general preference for the right over the left, that one should first strap on the right sandal, then the left. Rabbi Yohanan rebukes him: "Babylonian, don't do so, for the early ones didn't do so. Rather, when one binds, one binds the left, then the right, so that the right doesn't appear injured." In Rabbi Yohanan's view, sandals are worn chiefly to protect the feet. Hence one might sometimes go barefoot, if both feet are uninjured and in no special need of protection. One might also walk about with a sandal on one foot only, if that foot alone is injured. If one chooses to bind sandals on both feet, someone witnessing the beginning of the procedure could provisionally infer, from the placement of the first sandal, that the individual was attempting to protect an injured foot. The strapping of the second sandal would refute this inference. Nevertheless, because the inference can always arise, and because it is better for onlookers to suppose that one's left foot rather than one's right foot is injured, the "early ones" strapped on their left sandals before their right. It is uncertain whether Rabbi Yohanan provides the correct explanation of the custom of the early

[9] On the virtuous judge see Solum 2003.

ones.[10] What is important for our purposes is that Rabbi Yohanan takes the early ones as his model, and that the Talmud, in turn, offers him to us as one who imitates the early ones. Elsewhere, too (yBerakot 7:3 [11c]), on another matter of etiquette concerning left and right—whether, when holding in one's right hand the cup of wine on which one will recite a blessing, one may also support the cup with one's left hand—Rabbi Yohanan cites the early ones as authorities.

The close connection between etiquette and exemplarity is evident from a common reflex in works of scriptural exegesis from the earliest (tannaitic) stratum of rabbinic literature. Some of these works consistently derive rules of etiquette from the actions of biblical figures. Thus, for example, in Exodus 19, God proposes to Moses a covenant with Israel (19:3–6). Moses informs the elders of the terms (19:7), the people assent to the covenant (19:8a), and Moses reports back to God with their response (19:8b). A rabbinic commentary (Mek. R. Ishmael *ba-hodesh* 2) on Exodus 19:7 and Exodus 19:8b runs:

> "And Moses came and called to the elders of the people": It tells that Moses gave honor to the elders. . . . "[And placed before them] all of these words": The first, first, and the last, last. . . . "And Moses reported back the words of the people to the Lord": And did Moses need to report back? But the Torah taught etiquette. Moses went and reported back to him who sent him, for so Moses said: Even though he knows and witnesses, let me report back to him who sent me.

The above text constructs Moses as the consummate ambassador, according due respect to Israel's leadership, conveying the message in the correct order, and returning back with Israel's reply. The rabbinic text marks this last element of his performance as pedagogical ("The Torah taught etiquette") precisely in light of its superfluity: Moses reported back to God not because God was ignorant of Israel's reply, but because it is good form.

God, too, can serve as a model of proper etiquette. A tractate on etiquette from around the end of the classical rabbinic period (*Derek Eres Rabbah* 5:3) offers: "A man should not enter his friend's home suddenly. And let everyone learn etiquette from the Place (God), who stood at the entrance to the garden and called to Adam, as it says: 'And the Lord God called to Adam and said to him: Where are you?' (Genesis 3:9)." Here, as in the previous example,

---

[10] In the discussion that attends the Bavli's version (bShabbat 61a), it appears that the reason for binding the left foot first is the analogy to phylacteries, which are bound on the left hand. (Rabbi Yohanan does observe, even in this version, that binding the sandal on the right foot implies that it is injured, but he does not identify this implication as the reason for binding the left sandal first.) Later in the same context (bShabbat 66b), Abaye reports on his mother's authority that all binding should occur on the left. One gets the impression from these statements and others (e.g., Avin's, ibid.) that the sandal, because it was strapped, was partially assimilated to the category of magical amulets, which are characteristically fastened by straps.

but in a different way, God's omniscience makes the act in question superfluous and hence especially instructive. God called out to Adam before entering the garden not because he was unsure of Adam's whereabouts, but because it is polite to knock before entering. The tractate proceeds immediately to a story that incorporates this lesson into an almost comically detailed paean to etiquette.[11]

> A story: Four elders were on their way to attend a private counsel [in Rome], and they had an acquaintance by the name of Philosophos. And they were: Rabban Gamaliel, Rabbi Joshua, Rabbi Eleazar son of Azarya, and Rabbi Akiva. Said Rabbi Joshua to Rabban Gamaliel: Rabbi, is it your wish that we go and receive Philosophos our friend? He said to him: No. The next morning he said to him: Rabbi, shall we go and receive Philosophos our friend? He said to him: Yes. Rabbi Joshua went and knocked on the door. Philosophos considered to himself: This can only be the etiquette of a sage. [He heard the knock] a second time, and stood and washed his hands and feet. A third time, and he stood and opened the door. And he saw the sages of Israel, with these coming from here, and these coming from there, in this fashion: Rabban Gamaliel in the middle, Rabbi Joshua and Rabbi Eleazar son of Azarya to his right, and Rabbi Akiva to his left. Philosophos considered to himself: How shall I greet the sages of Israel? If I say: Peace be upon you, Rabban Gamaliel, I will slight the sages of Israel. And if I say: Peace be upon you, sages of Israel, I will slight Rabban Gamaliel. Once they reached him, he said: Peace be upon you, sages of Israel, and Rabban Gamaliel at the head. And not from Rabban Gamaliel do we learn [that the individual with the highest stature takes the middle position], but from the Presence we learn, for it says: "I saw the Lord sitting on his throne, and the whole host of heaven standing over him to his right and to his left (1 Kings 22:19)."

Every action in the story is undertaken with careful attentiveness to questions of propriety: how to address one's superior, when to call upon an acquaintance, how to make one's presence known, how individuals of different stature should arrange themselves in space, how to greet a group composed of individuals of high but varying stature.

Interest in the relationship between etiquette and exemplarity is not confined to the classical rabbinic corpus from late antiquity. Indeed, it becomes manifest at a programmatic level in the medieval code of law produced by Maimonides, the *Mishneh Torah*.[12] In the first three chapters of his "rules of character traits" (*Hilkot De'ot*) he explicates his Aristotelian doctrine of the mean (chapter one), details exceptions (chapter two), and then attempts (chapter three) to divert the

[11] See Loopik 1991: 102: "One gets the impression that the entire story was formulated especially as an example of a number of norms of etiquette." My translation of the passage depends in part on Loopik's.

[12] On this work, and its place in Maimonides's oeuvre, see, most recently, Halbertal 2014: 164–96.

impulse toward ascetic extremes from practice ("I will eat no meat") to intent ("I will eat meat, but only because it makes my body healthy and hence allows me to attend to non-bodily things"). In chapter four, Maimonides, putting on his doctor hat, compiles rules for healthy living, on the ground that "it is impossible to understand or know anything about the creator when one is sick" (4:1). The introduction to chapter five (5:1) identifies its topic as etiquette.

> Just as the sage is recognizable by his wisdom and his character traits, and thus distinguished from the rest of the crowd, so must he be recognizable by his actions: by the way he eats, the way he drinks, the way he has intercourse, the way he uses the bathroom, the way he talks, the way he walks, the way he dresses, the way he manages his affairs, and the way he conducts business.[13]

The remainder of the chapter takes up each category in turn, from eating to business practices. Here, for example, are some of the rules pertaining to talking and dressing (5:7, 9).

> A sage should not shout or screech like an animal when he speaks. Nor should he raise his voice excessively. He should rather speak pleasantly with everyone. If he sees that his speech is effective, and heeded, he speaks, and if not, he is silent. . . . Nor should he repeat himself, or speak too much, or too little. . . . A sage's clothing should be attractive and clean, and it is forbidden for a stain or grease or the like to be found on them. And he should not wear kingly clothing, as of gold or purple, at which everyone would stare, or a pauper's clothing, which disgraces the wearer, but rather average, attractive clothing. And his skin should not be visible beneath his garments, as with the very light flax material produced in Egypt.

These rules, like almost all of the others in the chapter, construct the sage as the quintessentially dignified individual. A civilized man, he neither sounds like an animal nor shows his skin like an animal. Nor, however, does he put on airs: his state is sagely, not kingly.

To appreciate the importance of the discussion of etiquette in chapter 5 of the *Mishneh Torah*, we must reflect briefly on the structure of Maimonides's code. Each section of the code is organized around the commandments of the Torah that pertain to it. The section on the "rules of character traits" includes eleven commandments, all but two of which are described in rapid succession in the final two chapters, beginning at 6:3.[14] The first, given at 1:5–6, is to

---

[13] Maimonides depends in part on Sifre Deuteronomy 343: "Sages are recognizable by the way they talk, the way they walk, and the way they wrap themselves."

[14] The two commandments analyzed here are positive commandments six and eight in Maimonides's *Book of Commandments*.

"imitate" God by cultivating such traits as kindness. The second occurs at 6:2, and from its position and content, Maimonides seems clearly to offer it as a corollary of the rules detailed in chapter 5.

> There is a positive commandment to cling to the sages and to their students, so as to learn from their actions, as it says: "And cling to him" (Deuteronomy 10:20). But is it possible for a person to cling to the Presence? Rather, our sages said in explaining this commandment: Cling to the sages and their students. Hence a person must make efforts to marry the daughter of a sage, and marry off his daughter to a sage, and eat and drink with sages, and do business with them, and attach himself to them in all sorts of ways, as it says: "And to cling to him" (Deuteronomy 30:20).

The rabbinic passage on which Maimonides relies ("our sages said") for his interpretation of the biblical injunctions to cling to God probably meant, in origin, something very different. Maimonides alludes to the following passage, from the Babylonian Talmud (bKetubot 111b).

> "To love the Lord your God and to cling to him" (Deuteronomy 30:20). But is it possible to cling to the Presence? Rather, one who marries off his daughter to a sage, and who does business with sages, and who bestows benefit upon sages from his property, is deemed by Scripture to cling to the Presence.

The great Talmud commentator, Rashi, interprets the passage, no doubt correctly, as a call for material support of sages. Maimonides instead adapts it so that it commends attaching oneself to sages in order to imitate them.[15] Thus, for example, conducting business with them becomes a means, not of underwriting their study, but of imitating their business practices. Maimonides likewise introduces the notion of eating and drinking with sages, to echo his reflections in chapter 5 on the sagely etiquette of eating and drinking. For Maimonides, then, as God serves as ethical exemplar, the sage is the model of etiquette.

Maimonides is far from the only medieval Jewish thinker to foreground such a role for the sage. The aforementioned Rashi, in weighing two different interpretations of a talmudic passage (bEruvin 54a s.v. *she-masrehet*), disfavors the one that would valorize the sage who is not particular about eating only fully cooked food, because in fact "we find in all of the Talmud that the way of sages is to be punctilious about their food as a matter of stature and honor, and about their clothing, as it is written, 'All who hate me love death' (Proverbs 8:36)." Rashi's use of Proverbs 8:36 depends on the interpretation of this verse elsewhere in the Talmud (bShabbat 114a), according to which it hyperbolically

[15] Eve Krakowski points out to me that echoes of Maimonides's revision may perhaps be found in a letter by his son Abraham in which he urges his addressee, a judge, to accept the proposal of one who would marry his daughter. See Goitein 1964: 196.

condemns to death the sage who, by wearing a garment stained with grease, "causes me to be hated," that is, gives God a bad name. The same talmudic passage underlies Maimonides's determination that a sage is forbidden from wearing a stained garment.

3. Etiquette as a Normative Form

The pervasiveness of exemplarity in Jewish sources on etiquette becomes relevant for the problem of law and exemplarity (or law and character) in Judaism when we consider the distinctive structural profile of etiquette as a normative form. In particular, etiquette integrates features of ethics and features of law. To summarize briefly what we will flesh out later: a number of philosophers defend the view that etiquette encodes important virtues, and that acting in accordance with etiquette, or some aspect thereof, is importantly virtuous. Etiquette resembles law, on the other hand, insofar as it is intrinsically determined by rules and conventions. My thesis is that, as a law-like domain that nevertheless allows for exemplarity, etiquette enables rabbinic Judaism to make sense of the notion of embodied law, or law that is both law in the ordinary, heteronomous sense, but also an agent of character formation.

A. ETIQUETTE AND LAW IN JEWISH SOURCES

I preface the philosophical analysis of etiquette with Jewish sources that highlight the continuity between etiquette and law, and that incidentally offer further insight into the role of exemplarity in the realm of etiquette. In the preceding passage about sagely dress, Maimonides slips into the language of law: "[I]t is forbidden for a stain or grease or the like to be found on [the sage's clothing]." The legal category of prohibition ("forbidden") recurs in a story about bathroom etiquette from the Palestinian Talmud (yShabbat 3:3 [6a]). Rabbi Joshua ben Levi introduces the proposition that, while discussions about the law may not ordinarily occur in a bathhouse or in a bathroom, one may inquire in these venues into laws that govern behavior there. After citing a story in support of this view, the Talmud challenges it with a different story. Samuel asked Rav, evidently when they were in a bathhouse or a bathroom, whether one may answer *amen* in a "dirty place" to a blessing. (Presumably, the blessing was recited in a clean place, but loud enough to be heard elsewhere.) Rav answers: "It is forbidden, and the 'forbidden' that I said to you is forbidden." Rav is self-consciously paradoxical. In demarcating a prohibition against religious speech (the *amen*) in the bathroom, Rav himself utters religious speech (a legal ruling) in the bathroom, and so violates his own prohibition. Indeed, he violates the prohibition twice, once in forbidding the *amen*, and again in forbidding discussion of the *amen*. The category of "forbidden" is ambiguous in Rav's usage.

Though it comes from the field of law, it cannot here carry real legal force, for Rav would not knowingly and willingly have violated the law. His rule, though expressed in legal terms, instead has the force of etiquette. Or in other words, Rav, by highlighting the fact that he violates the rule in the very act of articulating it, weakens its normative force. The story dramatizes the difference between two aspects of rabbinic authority in the realm of etiquette, namely, teaching and modeling. To teach proper etiquette, Rav must act improperly.[16]

Like the deployment of the word "forbidden" by Maimonides and by Rav, but in a different way, liturgical practices straddle the boundary between law and etiquette. A text from the Bavli (bBerakot 34b), for example, addresses the question of bowing during one of the blessings of the daily prayer (the *Amidah*). The Talmud reports that Rava bowed at the beginning of the blessing and at its end. The rabbis saw him and questioned him, on the assumption that one should bow, if at all, only at the beginning of the blessing. Rava cites precedent: "I saw Rav Nahman bow [twice], and I saw Rav Sheshet do the same." Confronted with an authoritative text that appears to contradict his practice, he (or the Babylonian Talmud speaking for him) interprets it so that it does not. What is conceptually basic, what precedes and determines interpretation, is the behavior of Rava's rabbinic predecessors. In any number of similar cases (e.g., of Rav Ashi in bBerakot 30a; of Rabbi Abba ben Zimna in yBerakot 6:1 [10b]), rabbis determine correct liturgical practice on the basis of nothing other than what they have seen other rabbis do. While liturgy is unquestionably a category of ritual law, its foundation lies in etiquette, where the governing question is how one ought to appear and act in the presence of a superior, God.[17] The connection between liturgy and etiquette becomes explicit in, for example, Rabbi Judah the patriarch's assertion (bBerakot 50a) that one who in the grace after meals praises God "*from* whose goodness we live" rather than "*through* whose goodness we live" betrays a lack of cultivation.[18]

The above sources underscore the relationship between etiquette and Jewish law. In the case of reciting *amen* in the bathroom, a rule of etiquette employs, if in a non-technical way, the terminology of law. The rules governing liturgical practice, which are as "legal" as any other body of ritual law, encode conceptions of proper etiquette vis-à-vis God. Having positioned etiquette between exemplarity (in section 2, "Exemplarity and Etiquette in Judaism") and law (this section) in Jewish sources, we now turn to philosophical treatments of etiquette that do the same.

---

[16] On apparently improper behavior by exemplars in the Confucian context, see Olberding 2012: 121–135.

[17] On this point see, especially, Ehrlich 2004a.

[18] The difference in Hebrew is but a single letter. Cf. bBerakot 64a, which adjudicates between two parting wishes that likewise differ by only a single letter marking the preposition. On this passage see Ehrlich 2004b: 19. These differences need not, at least in origin, have conveyed differences in meaning. See Haber 2009. See also Septimus 2011; tMegillah 3:27.

B. ETIQUETTE BETWEEN LAW AND ETHICS

Robert Pippin, in his analysis of civility, gives expression to the intuition that the normative space defined by etiquette borders on and overlaps with those of law and ethics. He sets out three classroom scenarios: A student steals another student's purse, or plagiarizes her work, or sighs loudly at her contributions to the discussion. The first offense is illegal, the second is legal but immoral, and the third is not immoral but is uncivil.[19] The ordering of the scenarios conveys the sense that, in terms of force, on the one hand, law and ethics represent the most serious normative frameworks, while etiquette comes in a distant third. With respect to form, on the other hand, civility is, as Pippin observes, the kind of thing that can at least in principle be legislated, and is in the case of political correctness laws. Political correctness laws tend to inspire "uneasiness," but their existence testifies to a basic homology—a formal likeness—between civility and law.[20] Judith Martin ("Miss Manners") similarly imagines law and etiquette complementing each other by encapsulating moral principles in the form of rules, where law takes on the more weighty principles, and etiquette the less important ones. "Civilized societies must use etiquette as a second system of restraint, so that they do not have to call in the mighty law every time someone does something disgusting."[21]

Etiquette is thus law-like first and foremost because it is constituted by society-specific rules. The crucial difference between the rules of etiquette and legal rules, and what ultimately makes obedience to the rules of etiquette, much more than obedience to legal rules, a saliently moral course of action, is that the rules of etiquette constitutively communicate, or express, moral commitments. To follow a good law, e.g., prohibiting theft, is to act morally, but not (except incidentally) to express a moral commitment. To act politely, by contrast, is to express respect. Cheshire Calhoun thus remarks of civility: "[W]hat makes being civil different from being respectful, considerate, or tolerant, is that civility always involves a display of respect, tolerance, or considerateness. Thus civility is an essentially *communicative* form of moral conduct. . . . [B]ecause communicating our moral attitudes is central to civility, being genuinely civil—unlike, say, being genuinely considerate or genuinely tolerant—requires that we follow whatever the socially established norms are for showing people considerateness, tolerance, or respect."[22]

---

[19] Pippin 2000: 104–105.

[20] Ibid.: 109. Another contribution in the same volume, Cahoone 2000, makes other helpful observations about the relationship between civility and law.

[21] Martin 1993: 351.

[22] Calhoun 2000: 260 (italics in the original). On such elements of etiquette as the "elaborate Victorian fork rules" that appear to encode nothing other than "class distinction," see ibid.: 251. See also Buss 1999.

Likewise, in a recent treatment of the concept of dignity, Michael Rosen defends the importance of dignified behavior—which in his analysis amounts more or less to the same thing as adherence to the rules of etiquette—by observing that "[t]he dignified person . . . is someone who *shows* dignity in their character or bearing . . .. [I]t is a manifestation of a moral capacity—the ability to resist natural impulses."[23] Being dignified is, fundamentally, "something that one shows."[24] As a form of display, being dignified is intimately related to treating others with dignity, understood as something that "requires that one *shows* [*sic*] them respect."[25] The concept of dignity, on this approach, trucks in the "symbolic or expressive."[26] To act in a dignified manner, and even to respect others' dignity, is a duty to oneself (as well as, at least in the latter case, a duty to others) that manifests itself in meaning-bearing, interpretable actions.[27] The expressive element of the action can figure less or more importantly. When, for example, I save another person from drowning, I convey respect for the drowning person's dignity, but the more relevant moral fact is, of course, that I have saved his life.[28] In other cases, the expressive element plays a much more prominent role. Rosen thinks of situations in which "we consider it right to treat people in ways that go drastically against the ways in which they themselves wish to be treated," as in war and criminal punishment. Here, even as we intentionally harm, we take care to avoid humiliation or degradation. And it is in such cases that, as with civility, conventions enter: "What *counts* as humiliation or degrading treatment varies drastically from culture to culture."[29] Convention is still more determinative in the case of acting dignified, the bread and butter of etiquette. Thus, for example, although dignified behavior cross-culturally entails, among other things, refraining from animal-like actions, what precisely such actions are depends on the particular culture.[30]

Karen Stohr persuasively argues that, because etiquette is constitutively expressive, one cannot be truly well-mannered unless one is also moral. To act politely toward another person is to express a minimal measure of respect for him. If one acts politely not because one respects (or knows that one ought to respect) him, but because one wishes, say, to win a contract that he controls, one is not in fact acting politely.[31] What we may call the motivational transparency of etiquette—its dependence on motivation—distinguishes it from law,

[23] Rosen 2012: 57. The italics here and in the other quotations from this book are the author's.
[24] Ibid.: 58.
[25] Ibid.
[26] Ibid.: 126.
[27] On duties to oneself, see ibid.: 142–156.
[28] Ibid.: 126–127.
[29] Ibid.: 127.
[30] On dignity and the human-animal boundary, see ibid.: 159–160.
[31] Stohr 2006: 196–199.

and explains, in part, why the field of etiquette, unlike that of law, allows for exemplarity.

A second reason that etiquette embraces exemplarity is that, despite the fact that it expresses itself in rules, the proper application of the rules of etiquette demands, again per Stohr, considerable practical wisdom, because the rules are rife with limitations.[32] The limitations are of two related sorts. First, because the moral stakes involved in etiquette are relatively low, rules of etiquette are often outweighed by other considerations. Stohr draws an example from Jane Austen's *Pride and Prejudice*. Elizabeth Bennet rightly undertakes a long walk, alone, to visit her sister, who has taken ill at a third party's house, even though it is unseemly for a woman to walk such a distance alone, and even though the journey dirties her dress beyond the point of presentability.[33] The second sort of limitation involves not violation of etiquette but doing more than what etiquette demands. Because etiquette manifests only *mere* respect, its rules must be dispensed with in situations that call for something more than respect. To send one's sister a note of congratulations upon her graduation from college would, ordinarily, be wrong, because one ought instead to respond much more effusively. Because obeying a law is not an expressive act, and because laws typically involve relatively higher-stakes questions and therefore are subject to fewer exceptions, analogous limitations figure more rarely in the case of law. Adherence to the rules of etiquette thus demands practical wisdom to a considerably greater degree than does abiding by the law.

4. The Solution

Like law, etiquette is a rule-based system that encodes moral commitments. But observance of the rules of etiquette demands engagement with the moral commitments that they encode to a degree, and with a directness and transparency, unknown to the practice of law-following. Etiquette thus occupies a unique position between law and ethics. This in-between status of etiquette, I suggest, allows it to play an important and heretofore underappreciated role within the normative structure of Judaism.

I argued in the first section of this paper that Judaism wishes to hold on to a paradox. It insists, on the one hand, that *halakhah* (Jewish law) is law in a real sense, i.e., a set of rules commanded by an authority, and not, as in the view of the allegorists whom Philo condemns, merely a thinly veiled collection of ethical principles. Because Judaism construes *halakhah* legalistically, it must concede Ze'ev Maghen's point: *Halakhah* is meant to police from the outside. But

[32] Ibid.: 202–211.
[33] Ibid.: 200–201.

it also wants to claim that *halakhah* plays a role in character formation. The fact that *halakhah* encodes virtues addresses the second horn of the dilemma, but to go too far in this direction is to confront the first horn. This is where etiquette enters. Insofar as etiquette is law-like but also implicates character, it represents a normative space in which law, not as such, but as a rule-based normative form, can become embodied. If in the realm of law Judaism can only tacitly and anxiously advance the claim that rule-following shapes character, then the realm of etiquette furnishes a context in which it can do so more explicitly and safely.

The intuitive expression of this argument is that when a person educated in the Jewish tradition acts in a well-mannered fashion, the notion that his character has been shaped by *halakhah* becomes more compelling. Civil behavior has this effect to a greater extent than moral behavior per se (e.g., informing a seller that he is mistakenly demanding less than the list price), because the former is bound by its form (i.e., rule-following) to law in a way that the latter is not. In short, in the well-mannered person, rule-following and, by extension, law find expression as grace.[34]

References

Buss, Sarah. (1999). "Appearing Respectful: The Moral Significance of Manners," *Ethics* 109: 795–826.

Cahoone, Lawrence. (2000). "Civic Meetings, Cultural Meetings." In *Civility*, ed. Leroy S. Rouner, pp. 40–64. Notre Dame, IN: Notre Dame University Press.

Calhoun, Cheshire. (2000). "The Virtue of Civility." *Philosophy and Public Affairs* 29: 251–275.

Collins, John J. (1972). *Jewish Wisdom in the Hellenistic Age*. Louisville, KY: Westminster John Knox Press.

Colson, F. H., and Whitaker, G. H. (1939). *Philo, Migration of Abraham*. Loeb Classical Library 261. Cambridge, MA: Harvard University Press.

Duff, R. A. (2006). "The Virtues and Vices of Virtue Jurisprudence." In *Values and Virtues: Aristotelianism in Contemporary Ethics*, ed. Timothy Chappell, pp. 90–104. Oxford: Oxford University Press.

Ehrlich, Uri. (2004a). *The Nonverbal Language of Prayer: A New Approach to Jewish Liturgy*. Translated by Dena Ordan. Tübingen: Mohr Siebeck.

Ehrlich, Uri. (2004b). "Taking Leave and Giving Leave: A Chapter in the Rules of Etiquette." In *Shefa' Tal: Studies in Jewish Thought and Culture (Bracha Zak Festschrift)*, ed. Z. Gries et al., pp. 13–26. Beer-Sheba: Ben-Gurion University Press

[34] My thanks to James Redfield for his incisive comments on parts of this paper; to the editors, Christian Miller and Angela Knobel, for their sagacious suggestions; and to the Character Project at Wake Forest University and the John Templeton Foundation for their grant support. The views expressed herein are mine alone, and do not necessarily reflect those of The Character Project, Wake Forest University, or the John Templeton Foundation.

Goitein, S. D. (1964). "Documents on Abraham Maimonides and His Pietist Circle." *Tarbiz* 33: 181–197.

Haber, E. (2009). "Significance of the Interchanging of the Mono-Consonantal Prepositions in Biblical Hebrew: Rabbinic and Comparative Semitic Evidence." *Maarav* 16: 29–37.

Halbertal, Moshe (2014). *Maimonides: Life and Thought*. Princeton, NJ: Princeton University Press.

Heinemann, Isaac (1953/2008). *The Reasons for the Commandments in Jewish Thought: From the Bible to the Renaissance*. Translated by Leonard Levin. Boston: Academic Studies Press.

Loopik, Marcus van. (1991). *The Ways of the Sages and the Way of the World: The Minor Tractates of the Babylonian Talmud:* Derekh 'Eretz Rabbah, Derekh 'Eretz Zuta, Pereq Ha-Shalom. Tübingen: Mohr Siebeck.

Maghen, Ze'ev. (2011). "Dancing in Chains: The Baffling Coexistence of Legalism and Exuberance in Judaic and Islamic Tradition." In *Judaic Sources and Western Thought: Jerusalem's Enduring Presence*, ed. Jonathan Jacobs, pp. 217–237. Oxford: Oxford University Press.

Martin, Judith. (1993). "A Philosophy of Etiquette." *Proceedings of the American Philosophical Society* 137: 350–356.

Mittleman, Alan. (2012). *A Short History of Jewish Ethics: Conduct and Character in the Context of Covenant*. Malden, MA: Wiley-Blackwell.

Niehoff, Maren R. (2011). *Jewish Exegesis and Homeric Scholarship in Alexandria*. Cambridge: Cambridge University Press.

Niehoff, Maren R. (2013). "A Jewish Critique of Christianity from Second-Century Alexandria: Revisiting the Jew Mentioned in Contra Celsum." *Journal of Early Christian Studies* 21: 151–175.

Olberding, Amy. (2012). *Moral Exemplars in the* Analects: *The Good Person is* That. New York: Routledge.

Pippin, Robert. (2000). "The Ethical Status of Civility." In *Civility*, ed. Leroy S. Rouner, 223–238. Notre Dame, IN: Notre Dame University Press

Rosen, Michael. (2012). *Dignity: Its History and Meaning*. Cambridge, MA: Harvard University Press.

Septimus, B. (2011). "Nevarek (l)elohenu: Linguistic Purity and the History of Halakhah." In *Ta Shma: Studies in Judaica in Memory of Israel M. Ta-Shma,* ed. Avraham (Rami) Reiner et al., 2.565–580. Alon Shevut: Tevunot Press.

Solum, Lawrence B. (2003). "Virtue Jurisprudence: A Virtue-Centered Theory of Judging." *Metaphilosophy* 34: 178–213.

Stohr, Karen. (2006). "Manners, Morals, and Practical Wisdom." In *Values and Virtues: Aristotelianism in Contemporary Ethics*, ed. Timothy Chappell, 189–211. Oxford: Oxford University Press.

Weinfeld, Moshe. (1972). *Deuteronomy and the Deuteronomistic School*. Oxford: Clarendon Press.

24 }

Christian Character Formation and the Infusion of Grace

Ray S. Yeo

Introduction

Central to the thought of Jonathan Edwards is his theologically sophisticated psychology of grace, or what is sometimes called his "religious psychology of the heart.."[1] Discussion of this important subject can be found not only in his famous "Treatise Concerning Religious Affections" but also throughout his vast corpus of writings, including his sermons and miscellanies entries. Edwards's understanding of the psychological functioning of grace may not be entirely original; much of what he says can be found in some form in his Puritan and Protestant forbearers. However, his treatment of the subject is unsurpassed in the Puritan literature in terms of the level of precision, thoroughness, and systematic sophistication.[2] The heart of his account comes down to his foundational doctrine of the *infusion of grace*, which was meant in part to explicate the biblical idea of the Holy Spirit's indwelling reality within believers.

Meticulous and systematic as Edwards was, his theory of infusion did leave some important issues unaddressed, and it could benefit from the important work of the contemporary philosopher William Alston, who shares with Edwards a similar interest to develop a model for understanding the indwelling work of the Holy Spirit in the psychological lives of believers. This chapter is an attempt to retrieve and, with the aid of Alston's work, further develop a broadly Edwardsean view of Christian character formation by grace.

---

[1] Portions of this chapter have appeared in Ray S. Yeo, "Toward a Model of Indwelling: A Conversation with William Alston and Jonathan Edwards," *Journal of Analytic Theology* 2 (2014).

[2] Brad Walton came to this conclusion after surveying a wide range of Puritan writings. See Brad Walton, *Jonathan Edwards, Religious Affections and Puritan Analysis of True Piety, Spiritual Sensation and Heart Religion* (New York: Edwin Mellon Press, 2002), 231.

Before we begin the discussion, it should be noted that the purpose of the chapter is simply to explore one aspect of the indwelling of the Spirit that is of common interest to Alston and Edwards. It is not meant to be an exhaustive account for the reality of indwelling but simply a modest constructive proposal for understanding an important aspect of indwelling that draws upon the work of Alston and Edwards. Specifically, the focus is on the individual psychological transformation of believers.

The attention on the individual psychology of the believer is not meant to undermine the social and ecclesiastical dimension of indwelling in any way. The individual and social aspects of indwelling grace are both essential to the Christian life. Nevertheless, to echo Alston, despite the importance of the social dimension, it must also be recognized that the indwelling operation of the Holy Spirit in the process of Christian transformation does not "bypass the inner psychological development of each individual" either.[3] Hence, for the purpose of this chapter, the focus will simply be on the individual aspect of indwelling as highlighted by Edwards and Alston. With this aim in mind, we shall begin by examining Edwards's doctrine of the infusion of grace, followed by Alston's model of indwelling. The implications of the Edwards-Alston synthesis for Christian character formation are then considered.

Grace as Infusion

Grace, for Edwards, denotes not only the gracious free judicial justification by God for sinners but also the reality of the infusion of the Holy Spirit in the hearts of believers.[4] He makes an important distinction between the gift of "infused grace"[5] and other gracious gifts of the Spirit that do not involve infusion. The key difference between the two is that infused grace involves an influence of the Spirit that is internal and inherent within the subject's nature and heart whereas the influence of non-infusive grace remains external.[6] By this he means that the Spirit's work of infusion exerts a kind of inhering and abiding influence within the subject so that it becomes an integral aspect of the individual and the very nature of the individual can be rightly said to have changed into a spiritual nature.[7] In contrast, there are other more external

---

[3] William P. Alston, "The Indwelling of the Holy Spirit," in *Divine Nature and Human Language: Essays in Philosophical Theology* (Ithaca, NY: Cornell University Press, 1989), 225.

[4] *The Works of Jonathan Edwards, vol. 2, Religious Affections*, ed. by John E. Smith (New Haven, CT: Yale University Press, 1959), 469. Henceforth, I shall use the abbreviation *WJE* for *The Works of Jonathan Edwards*, followed by the volume number and page number(s).

[5] *WJE* 13:171.

[6] *WJE* 8:157.

[7] Within the context of Edwards's doctrine of infusion, the term "spiritual nature" is typically used to denote a human nature that has been infused with the divine Spirit and thereby takes on various qualities that reflect the nature of God's Spirit. This does not imply that the infused human subject

influences, gifts and works of the Spirit that do not involve such integral and inhering change of quality or disposition within an individual's nature.[8]

Another important difference between infused grace and other non-infusive operations of the Spirit on the human mind is that the former introduces a new supernatural *principle* to the functioning of the soul or human nature, whereas the latter does not.[9] Edwards, echoing his Reformed predecessors, held that this crucial difference is also what distinguishes "common grace" from "special" or "regenerative grace." For him, "common grace is only the assistance of natural principles" and is hence limited to assisting "the faculties of the soul to do that more fully, which they do by nature."[10] Special grace, on the other hand, infuses and introduces a wholly new and qualitatively distinct *supernatural* and *spiritual* principle to the functioning of the faculties of the soul. It lays a new inhering *supernatural* foundation within the soul such that the recipients of special grace are able to exercise their faculties in a way that is above their natural potential.[11]

Edwards takes the infusion of a new spiritual principle to occur in the human heart or will. He conceives this new principle to be a kind of new fundamental inclination implanted within the faculty of inclination or will. However, he does not mean to isolate the effect of infusion to merely a single faculty. Edwards sees infusion as affecting the whole human person, including the understanding, and he is able to affirm this because he views the faculties of understanding and will (or inclination) as functionally intermeshed.[12] The regenerated will (i.e., a will infused with the spiritual principle) with its newfound spiritual inclination "consents" to the religious ideas in the understanding and therefore alters the axiological appearance of those very ideas to the subject. As remarked by Morimoto, "[T]he infusion of grace cures the will and, through this cured will, transforms the person's understanding faculty, so that the convert may see the beauty and holiness of God."[13] Hence, for Edwards,

---

ceases to be human. Edwards sees the infusion of the Holy Spirit as constitutive of God's original design for human nature. Non-infused human nature, then, on Edwards's view, is seen as a kind of defective or non-optimal human nature.

[8] *WJE* 2:202–203.

[9] Paul Ramsey observes that "Edwards used the word 'principle' in the sense of the Latin *principium* or the Greek *arché*. The word 'principle' means a source or beginning or spring of disposition and action. But it also means the direction, shape, or contours of human hearts and lives, as in the root of our word 'archetype,' or the arché or formative power of Plato's ideas, such as justice or beauty, or triangularity." Paul Ramsey, "Editor's Introduction," in *Works of Jonathan Edwards, vol. 8, Ethical Writings*, ed. Paul Ramsey (New Haven, CT: Yale University Press, 1989), 16.

[10] *WJE* 18:155.

[11] *WJE* 17:410–411.

[12] *WJE* 1:272.

[13] Anri Morimoto, *Jonathan Edwards and the Catholic Vision of Salvation* (University Park: Pennsylvania State University Press, 1995), 16.

the infusion of grace in the will also serves as the basis for the illumination of the understanding in the form of a consenting delight toward the idea of God.

The infusion of grace brings about a new psychological reality for the regenerate (i.e., those who are infused with the Holy Spirit) and serves as the basis for the spiritual life.[14] In particular, it grants them a supernatural *spiritual sense*, not available to natural man, which allows believers to perceive the holiness of God in supreme delight. The spiritual sense does not merely allow them to see more of divine goodness compared to the natural man; rather, the saints are able to experience something that is wholly and qualitatively distinct from any kind of sensations of the natural mind.[15] It is akin to gaining a new sense, not unlike a blind man gaining the sense of sight.[16] This new *spiritual sense* is Edwards's well-known theory of spiritual perception which has received much attention in contemporary Edwardsean studies.[17]

Infusion as Indwelling Spirit

One of the most important features regarding Edwards's understanding of infusion is that he sees the Holy Spirit as the very thing that is infused within the hearts of believers. For him, the infusion of grace within the soul as an inhering or internal principle is equivalent to the indwelling of the Spirit within the saints. Hence, he asserts that "there is no other principle of grace in the soul than the very Holy Ghost dwelling in the soul and acting there as a vital principle.""[18] Edwards follows in the Augustinian theological tradition and equates the *love of God* with the Holy Spirit and sees the person of the Spirit as that which is infused in the saints.[19] On his understanding, the disposition of love that draws us to God is the very same mutual love that binds the Father and the Son—the Holy Spirit.[20] Hence, when believers are infused with the Holy Spirit in the process of sanctification (i.e., Christian transformation),

---

[14] *WJE* 8:157–158.

[15] *WJE* 2:262.

[16] Ibid.: 205.

[17] For some of the important contemporary secondary literature on Edward's theory of spiritual perception, see Terrence Erdt, *Jonathan Edwards: Art and the Sense of the Heart* (Amherst: University of Massachusetts Press, 1980); Sang Hyun Lee, *The Philosophical Theology of Jonathan Edwards* (Princeton, NJ: Princeton University Press, 1988); Michael J. McClymond, *Encounters with God: An Approach to the Theology of Jonathan Edwards* (New York: Oxford University Press, 1998); Perry Miller, "Jonathan Edwards and the Sense of the Heart," *Harvard Theological Review* 41, no. (1948); John E. Smith, "Editor's Introduction," in *The Works of Jonathan Edwards. Vol 2, Religious Affections*, ed. John E. Smith (New Haven: Yale University Press, 1959).

[18] *WJE* 21:196.

[19] Augustine made this identification in *De Trinitate*, VI.7.

[20] *WJE* 21:121.

they are simultaneously caught up in the eternal Trinitarian love within the Godhead and thereby participate in the Trinitarian life.[21]

One significant consequence of Edwards's conception of indwelling grace as just sketched is that it introduces an element of *deification* or *theosis* to his doctrine of infusion. That is, his account involves the human subject partaking in and being united with the divine nature and thereby sharing in, to a certain extent, some of the divine qualities of God. According to Edwards, the divine Spirit is given to indwell the regenerate as a new internal "principle of life and action." This indwelling of the Spirit changes "the nature of the soul" and brings about a "new nature" that is in some sense "divine," according to the divine principle of life that is infused within it. The new nature of the soul now "admits divine light" and is able to perceive the glory of divine realities.[22] The element of deification (or theosis) in Edwards's account is especially evident in his "Treatise on Grace" in which he equates the new infused principle of grace in the heart of believers with the principle of divine love. As mentioned, this divine principle is nothing other than the third person of the Trinity, the Holy Spirit. Thus he states that "holy and divine love dwells in their hearts, and is so united to human faculties that 'tis itself become a principle of new nature."[23] For Edwards, the saints partake of the divine nature because "the nature of the Spirit of God is divine love" and that very same divine nature or love in the Holy Trinity is communicated and infused as a divine principle in the hearts of the redeemed so that the new nature of the soul becomes "spiritual" and, by extension, divine.[24] Hence, as it turns out, Edwards's explication of the indwelling or infusion of the Spirit amounts to a kind of *theosis*, a quite literal and direct partaking of and union with the divine Spirit such that the new nature acquired through the infusion of the person of the Spirit is also in some sense the divine nature.[25]

Edwards's account of the infusion or indwelling of the Spirit does raise some questions. In particular, how are we to understand the union of the person of the Holy Spirit with one's psychological faculties? More precisely, in what way can a divine person, as opposed to merely the effect of his action or influence, be said to be so united with the faculties within the soul of man so

---

[21] *WJE* 8:132.

[22] *WJE* 13:462–463.

[23] *WJE* 21:194.

[24] Ibid.: 191–192.

[25] Edwards did not use the term *theosis* explicitly, but a great number of Edwardsean scholars, including Stephen Holmes, Michael McClymond, and William Danaher, have considered Edwards's view of sanctification as a version of theosis. See Stephen R. Holmes, *God of Grace and God of Glory: An Account of the Theology of Jonathan Edwards* (Edinburgh: T&T Clark, 2000), 58; William J. Danaher, *The Trinitarian Ethics of Jonathan Edwards* (Louisville, KY: Westminister John Knox Press, 2004), 42. Michael J. McClymond, "Salvation as Divinization: Jonathan Edwards, Gregory of Palamas and the Theological Uses of Neo-Platonism," in *Jonathan Edwards, Philosophical Theologian*, ed. Paul Helm and Oliver Crisp (Burlington, VT: Ashgate Publishing, 2003), 153.

as to form a new nature and intrinsic vital principle? Unfortunately, Edwards does not elaborate much further on the nature of this seemingly mysterious union beyond the assertion that the divine Spirit is united with the soul of man as a divine principle in the heart.[26] However, without saying something more regarding such a foundational concept, Edwards's account remains underdeveloped and unilluminating in this important aspect. In the following section, I wish to draw upon the work of William Alston to further develop Edwards's psychology of infused (or indwelling) grace. The purpose is to provide a model for understanding the nature of the Spirit's indwelling union and presence in the human psyche as an inhering principle or disposition of love that was lacking in Edwards's original account. Hence, it is to Alston that we now turn.

Alston's Model of Indwelling

In his insightful essay, "The Indwelling of the Holy Spirit," Alston examines two common models for understanding the nature of the Spirit's work of transforming and sanctifying believers.[27] Beginning with the "fiat" model, Alston characterizes this position as one in which God directly produces "new dispositions and tendencies" within the psychological structures of believers and thereby transforms them into the kind of person he desires. It is a fairly one-sided affair with God simply implanting a new disposition within a person with minimal input, consent, and contribution from the human subject. God simply decrees the transformation and the individual is changed. Alston, however, criticizes this picture of sanctification and indwelling as being too impersonal and hence fails to do justice to the rich interpersonal and familial description of the Holy Spirit's indwelling ministry in the New Testament (e.g., Rom. 8:15–16).[28]

The second model he examines is the "interpersonal" model in which the ministry of the Spirit is understood more along the paradigm of interpersonal relationships and interactions. On this view, the Spirit sanctifies the believer in a manner similar to the way human beings usually influence one another, "seeking to evoke responses, voluntary and otherwise from the other person." Whereas on the "fiat" model, sanctification could in principle be accomplished without any kind of personal engagement between God and man, this is not possible on the "interpersonal" model. God transforms us through interacting

[26] Robert Caldwell came to the same conclusion noting that Edwards did not know how the saints' mysterious union with God works, but "that it does happen, Edwards was completely convinced." Robert W. Caldwell III, *Communion in the Spirit: The Holy Spirit as the Bond of Union in the Theology of Jonathan Edwards* (*Milton Keynes*: Paternoster, 2006), 120.

[27] Alston, "Indwelling."

[28] Ibid.: 233–235.

with us on an interpersonal level; it is an interaction that respects our "personal integrity," choices, and agency.[29]

Alston is more sympathetic toward the interpersonal model even though he finds the fiat model somewhat plausible as well, despite his initial criticism. Nevertheless, he argues that the two models ultimately fall short in one important aspect—they both fail to account for the "internality" of the indwelling ministry of the Spirit. Alston notes that the description of the Spirit's indwelling in the New Testament is a form of presence within believers that is over and above the Spirit's omnipresence. God is internally present to the saints in a way he is not to creation in general and none of the models examined is able to adequately account for this "special mode of internality."[30]

In light of the weakness Alston sees in the two models, he then argues for the "partial life sharing" account as a model for understanding the nature of the Spirit's indwelling presence. Such a model is understood as the human subject partially sharing in various psychological elements of the Holy Spirit, thereby resulting in a partial merging of psychological lives between the subject and God. In our everyday relationships with others, we do not have direct experiential access to the thoughts, beliefs, experiences, and motivations of another in the same direct manner that we have of our own, no matter how intimate a relationship may be. There are "physical and psychological barriers" between us that "insulate" one human psyche from another to some degree. A "partial life sharing" between the Holy Spirit and an individual would then involve a "breaking down" of such "barriers that normally separate one life from another," allowing believers to be more intimately related to God than to any other human being.[31] On this model, the Holy Spirit would share elements of his psychological life with the saints so that the elements might be "immediately available" to believers, just as their own psychological elements are available to them. Moreover, these shared psychological elements would also exert a certain level of influence on an individual's motivation, action, and cognition, just as one is typically influenced by his or her own psychological states and dispositions.[32]

The psychological elements that are shared may include experiences, feelings, attitudes, tendencies, values, or beliefs. The nature of such sharing is twofold. First, it is a cognitive sort of sharing. The believer shares in the divine psychological life by being cognitively aware of these psychological states in the same way that he or she is directly aware of his or her own. Second, it may also be a conative kind of sharing. The human subject is not only aware of the various psychological elements but also shares in certain divine conative

---

[29] Ibid.: 236–238.
[30] Ibid.: 241.
[31] Ibid.: 246.
[32] Ibid.

tendencies, such as love, compassion, and various motivations of the divine mind. This conative sort of sharing would allow individuals to affectively experience the divine tendencies, urges, or motivations in varying degrees of strength. According to Alston, the human agent's conative system need not simply be replaced by God's. Rather, he suggests that the divine conative tendencies may be introduced into the human agent's conative system in an initially "weak, isolated and fragile" manner, existing side by side with the agent's own sinful tendencies.[33] For Alston, it is possible that these new divine tendencies might be considered alien by the agent until he or she psychologically appropriates and nurtures them in a certain way.[34] The transition from the initially "weak, isolated and fragile" cognitive and conative sharing of the divine psychological elements to a "full-blooded" integration of these elements into one's psychological system may require a prolonged and drawn out process of responding freely and rightly to God's personal interactions and coming to own them in a certain way.[35]

Alston maintains that the cognitive and conative aspects of divine sharing constitute what it means to participate in the life of God. He contends that our spiritual rebirth could very well begin with a gracious initiative of God to introduce and share his divine psychological elements into our conative and cognitive systems. Although such introduction may commence in an initially faint and fragmented manner, it nevertheless provides the basic divine psychological resources from which they can mature organically through our free cooperation and, eventually, come to transform and sanctify us in comprehensive ways. In this fashion, believers literally partake of the divine psychological life and come to be united with God in a personal and internal manner.[36]

Alston's model is not meant to be taken as an exhaustive claim on the meaning of indwelling union with God. There might very well be other aspects of participation in the divine life that are not captured by his account. However, he thinks it is likely that indwelling union may at least include a sort of partial psychological union similar to what he has described.

The preceding account of the Holy Spirit's indwelling by Alston is quite insightful and turns out to be very complementary to Edwards's own understanding of Spirit infusion or indwelling. More significantly, it is able to fill out, in an illuminating way, Edwards's underdeveloped idea of an indwelling union between the person of the Holy Spirit and the human psyche. Therefore, I propose that Alston's *partial life sharing* model of indwelling can be fruitfully integrated with Edwards's theological psychology to provide an insightful account for understanding the nature of divinely infused grace that

---

[33] Ibid.: 251.
[34] Ibid.: 246.
[35] Ibid.: 249.
[36] Ibid.: 251–252.

is foundational to Christian character formation.[37] On this Edwards-Alston synthesized account, for a person to be infused with divine love is for the Holy Spirit to "break down" the "physical and psychological barriers" that normally separate one psychological life from another and "share," in the sense described by Alston, an aspect of his loving disposition or tendency with believers. This infused grace or love also serves as the foundational basis from which believers can begin to overcome the domination of sin and come to love the holiness of God.

Indwelling Grace and Christological Sonship

Having outlined a broadly Edwardsean account of indwelling grace, I would like, in the remainder of the chapter, to draw out its significance for practical Christian living and character formation. In particular, I want to focus on the idea of *formation in Christological sonship* that is central to the Christian New Testament writings.[38] At the heart of Christian character formation is the transformation of believers into the image of Jesus Christ, the eternal Son of God, as they are united with him and adopted into the family of God through sharing in his sonship.[39] Christian formation proceeds from this reality of adoption into the Sonship of Christ by which believers are able to relate appropriately to God as their Father.

The process of adoption in Christ is constituted by two central aspects. First, there is the conferring of the status of children, with its associated rights, privileges and responsibilities, to those who are redeemed by God. The second central aspect is the giving of the Spirit of God's only Son (i.e., Jesus Christ) to indwell believers so that they, too might enjoy the intimate emotional longing and connection that Jesus shares with God the Father as the natural (as opposed to adopted) and eternal Son of God.[40] This latter aspect of adoption is significant because while a person may be granted the status of an adopted child, he or she might not be able to enjoy the intimate emotional attachment and connection that natural children typically have with their parents. Hence, the way in which God seeks to overcome that affective alienation is to have the very Spirit of his natural Son indwell his adopted children.

---

[37] In a separate paper, I have argued that the Edwards-Alston synthesis encounters several difficulties and have provided a number of modifications to circumvent the problems raised. However, to pursue these issues in this chapter would detract from its overall aim. For the purpose of this chapter, I have not delved into the proposed modifications to the Edwards-Alston synthesis. For details on the modified account, see Yeo, "Toward a Model of Indwelling: A Conversation with William Alston and Jonathan Edwards," *Journal of Analytic Theology* 2 (2014)

[38] See, for instance, Galatians 4:5–6; John 1:12; Romans 8:15.

[39] See Romans 8:29.

[40] Cf. Galatians 4:6–7.

The broadly Edwardsean account of Spirit indwelling or infusion articulated earlier is very relevant to the idea of Christological adoption highlighted above. According to Edwards, the infused disposition that the indwelling Spirit brings about is also an aspect of the disposition of Jesus the Son to experience and enjoy the glory or goodness of God the Father. This is because Edwards sees, on the basis of various biblical allusions, that the Holy Spirit is also the Spirit of Jesus the Son.[41] Hence, in his sermon "The Excellency of Christ," Edwards asserts that the saint's enjoyment and joyous perception of the Father's goodness is in itself a partaking in the "Son's enjoyment of God." Moreover, he notes, the church, in being united with Christ "are in a sort partakers of his relation to the Father: they are not only sons of God by regeneration, but by a kind of communion in the Sonship of the eternal Son."[42] Hence, Edwards describes Christ as bridging the "infinite distance" between God and man through his incarnation to bring us into the bosom of the Father where he is, so that we may partake of the eternal enjoyment and communion of sonship that Christ uniquely enjoys with the Father.[43] For Edwards, this is as intimate an experience and perception of the Father's glory as creaturely beings can get. Thus, on the Edwardsean view, the infused spiritual disposition is also the unique disposition of Christological Sonship to affectively perceive and enjoy the goodness of God as Father in ways that are not possible without this disposition.

According to Edwards, this Christological disposition that is shared with believers through the indwelling Spirit consists of a certain motivation for union and a good-seeing disposition toward God. As mentioned, Edwards conceives of the infused principle as a new disposition or capacity to perceive and appreciate the goodness and beauty of God in joyful delight. This infused disposition is also a kind of love, since part of what it means to love someone is to appreciate the ways in which that person is good. Love is nourished and constituted by such an appreciation, without which it is difficult to see how anyone can love. In addition, Edwards also views the new disposition as playing an important unitive role in the believers' relationship of love with God. For him, love includes a unitive component; that is, it involves a longing to be together with the one loved and a coming to see the good of another as one's own. Hence, he argues that love "includes a desire of union with the object as well [as] a desire of the welfare of the object."[44] This unitive desire within believers is of course generated by the infused disposition, and it motivates one to "union and intercourse" with God.[45]

---

[41] *WJE* 2:199–200.
[42] *WJE* 19:593.
[43] Ibid.: 593–594.
[44] *WJE* 21:326.
[45] Ibid.: 326–327.

Drawing on Edwards's analysis, it is helpful to conceive of the Spirit-infused disposition as the result of the Holy Spirit's act of sharing Christ's unique good-seeing disposition toward God as Father with believers and the motivational drive to be united with him.[46] I shall term this dual aspect of the Christological disposition shared with God's adopted children, through the indwelling Spirit, as the *filial unitive drive* (or *unitive drive,* for short). Let us try to develop this notion a bit further.

A helpful way of conceiving the unitive drive is to see it as analogous to a human sex drive. A sex drive dawning in a young boy undergoing puberty may start off weak and without much representational content. He feels a mild urge toward something without knowing exactly what is it that he wants. It is only after acquiring various experiences, propositional content, and beliefs that he begins to come to understand clearly what he is drawn toward. Moreover, this growing sex drive also instills within him a newfound good-seeing disposition toward members of a certain gender and a certain added capacity to enjoy his relationships with them. Through the development of a sex drive he has come to acquire novel powers of appreciation, perception, and enjoyment of the ways in which they are good in a manner not possible before.[47] These new powers of his psyche in turn motivate him toward union of life with members of that gender.

In an analogous fashion, when the Holy Spirit indwells believers, he shares with them Christ's distinct drive toward union with God and his unique filial disposition to appreciate, perceive, and enjoy the goodness of his heavenly Father. The infused unitive drive hence opens up new motivations for union and life with God as Father that are unavailable and psychologically alien to those without this drive. In addition, like a slowly dawning sex drive, the sharing of the unitive drive with believers could also be gradual and need not entail an exhaustive sharing of that drive all at once. The filial disposition could

---

[46] The human loving tendency of Christ would also include his love for fellow humans. However, following Edwards, I take it that a truly virtuous love for neighbor is grounded in a love for God; hence, the sharing of Christ's human loving tendency is first and foremost the sharing of an aspect of his love for the Father. See *WJE* 8:142.

[47] Maurice Merleau-Ponty makes a similar point regarding the effect of a sexual drive or awareness in one's perception of the sensory world. In his discussion of a sexually impotent patient, he makes the following observation: "Perception has lost its erotic structure, both spatially and temporally. What has disappeared from the patient is his power of projecting before himself a sexual world, of putting himself in an erotic situation, or, once such a situation is stumbled upon, of maintaining it or following it through to complete satisfaction . . . At this stage one begins to suspect a mode of perception that is distinct from objective perception, a kind of significance distinct from intellectual significance, an intentionality which is not pure 'awareness of something.' Erotic perception is not a *cogitatio* which aims at a *cogitatum*; through one body it aims at another body, and takes place in the world, not in a consciousness." Merleau-Ponty, as cited by Mark Wynn. Mark Wynn, *Emotional Experience and Religious Understanding: Integrating Perception, Conception and Feeling* (New York: Cambridge University Press, 2005), 155.

be shared without sharing the full strength of that tendency instantaneously. Therefore, the experience of the unitive drive might very well begin as weak or partial and gradually grow only as believers mature spiritually.

Another significant parallel between a sex drive and the filial unitive drive, especially in the context of Edwards's theory of spiritual perception, is that they are both partially aesthetic in nature. An important aspect of the good-seeing disposition instilled by a sex drive is the capacity to perceive and appreciate the aesthetic beauty of another person in a manner quite distinct from perceptions of beauty without the sex drive. Our human sex drive brings about a unique kind of perception of sexual beauty that typically only arises in relation to perceptions of members of a certain gender. This sex drive is thus able to reveal a unique kind of sexually related aesthetic properties that can only be grasped through the sexual perceptions it affords. Similarly, the filial unitive drive grants one epistemic access to a distinct class of aesthetic properties regarding the beauty of God as Father that can only be perceived through the kind of perceptions the unitive drive makes possible.[48]

It should be noted that the filial unitive drive is not equivalent to a robust virtue of love for God that the Christian life demands. Thus, the sharing of this drive through the indwelling Spirit is not incompatible with the existence of sin and immaturity in the life of the regenerate. The unitive drive is simply an aspect of Christ's love and appreciation for the Father, and not his full virtue of love. Indeed, the love of God within Christ and the hearts of believers is not merely a bare good-seeing disposition or motivating drive; instead, it is a virtuous love that is constituted by concrete representational content, practical wisdom, and various skills of self-management that mere bare dispositions lack. Hence, it is important to make a distinction between the unitive drive that is infused and the resultant virtuous love of God that it gives birth to. The former in itself is a disposition without any clear representational content, understanding, and practical know-how, which is insufficient to be counted as a kind of psychological attitude or Christian virtue. The latter, on the other hand, is a robust and concrete virtuous love firmly oriented toward a flourishing communion with God the Father. Nonetheless, despite their differences, the unitive drive does serve as the basis for the formation of a virtuous filial love for God. Their relationship is analogous to that between a sex drive and virtuous romantic love. The dawning of a sexual drive introduces a new psychological disposition that in itself lacks any clear representational content, moral understanding, and practical know-how. Such a bare sexual disposition is not equivalent to romantic love, but it is foundational and constitutive

[48] Edwards conceives of this distinct class of divine aesthetic properties as equivalent to the beauty of holiness or the glory of God. For details on his thinking on divine aesthetics, see Roland Delattre, "Beauty and Theology: A Reappraisal of Jonathan Edwards," *Soundings: An Interdisciplinary Journal* 51, no. 1 (1968): 60–79.

of a whole range of romantic phenomena that comstitute virtuous romantic love, including romantic delights, appreciations, concerns, and attractions. In a parallel fashion, the infusion of the unitive drive, as Edwards puts it, lays a "new foundation" in our soul and allows us to use our human faculties in a new way.[49] It thereby opens up a whole new range of distinct emotions, experiences, motivations, concerns, and appreciations toward God that was not previously available.[50] Moreover, it also introduces a new spectrum of possible Christological filial virtues for the human subject that promote a flourishing filial relationship with God as Father.[51] In our concluding section, by drawing on the insights of Edwards's religious psychology, I would like to outline an account of one such important virtue in the Christian tradition—the virtue of Christological filial obedience.[52]

Infusion and Filial Obedience

Central to Edwards's psychology of Christian obedience is the idea of spiritual delight. He thinks that one important reason why many people, including self-professed Christians, continue to sin is that they find the pleasures of sin greater than the delight of obeying God. Edwards held that the commandments of God in themselves are not burdensome or difficult to fulfill but are only experienced as such because of the absence of a spiritual appetite within the unregenerate (i.e., those who are not infused with the Holy Spirit), resulting in a natural aversion to spiritual things.[53] Hence, even if they believe that the law of God is ultimately good for them, they cannot obey it wholeheartedly. This line of analysis and reasoning regarding the psychology and phenomenology of disobedience is quite consistent in Edwards's corpus of writings. It is present in his early sermon "True Love to God" and also in his later major works "Original Sin" and "The Freedom of the Will," in which he essentially attributes the "moral inability" of the unregenerate to choose and obey God to the absence of a fundamental kind of delight in the will.[54]

Edwards's solution to the problem of disobedience is the infusion of the love of God (i.e., the Holy Spirit) within the soul. In "True Love to God," he makes the following argument:

---

[49] *WJE* 2:206.

[50] *WJE* 17:411.

[51] Filial virtues refer to the virtues that relate to one's role as son or daughter. They would include such virtues as respect for one's parents and loyalty to one's family. In the context of the Christian faith, filial virtues refer to virtues that relate to one's role or status as an adopted child of God in Christ.

[52] The importance of filial obedience within the Christian faith can be seen in the biblical passage 1 John 5:1–5.

[53] *WJE* 10:636–637.

[54] *WJE* 3:381–383; *WJE* 1:196–198.

> The reason why they take the precepts of religion to be heavy and exceeding burdensome is because they find in themselves a great contrariety and aversion to it; [it] is directly contrary to their sin and corruption; but if ever the love of God, the contrary principle, is infused into their souls by the Spirit of God, their corruption will be mortified and the difficulty ceases with it.[55]

Edwards believes that the love for God that results from the infused divine love makes obedience to God "easy" because it is a delight to obey the one whom we sincerely love and most desire to please.[56] There is here a pleasure or delight that the regenerate experiences when they grasp that God, whom they love, is pleased in their act of obedience. This very pleasure is also what he later in the same sermon refers to as the "pleasures of communion with God."[57] And it is such delight that removes the psychological burden of God's commands, making them seemingly "easy and pleasant."[58]

Edwards's account of the psychology of Christian obedience via the infusion of grace is rather insightful for understanding the virtue of filial obedience. For the rest of the chapter, I hope to utilize his account, alongside the ideas developed earlier, to provide a brief sketch of the psychology of Christian filial obedience.

The infused unitive drive of Christological Sonship plays an important constitutive role in generating the pleasure or delight of filial communion with God as Father by providing the affective appreciation, attraction, and longing for the goodness of the Fatherhood of God. The unitive drive affectively infuses every aspect of the believers' interactions with God and generates for them a unique kind of filial delight in their positive interpersonal experience of the Father. As noted earlier, this is similar to the way in which our human sex drive, with its associated sexual attraction and longing, infuses every aspect of the experience of our interaction with those we are romantically drawn to, thereby generating a distinct romantic kind of emotional delight in our positive interpersonal experience with them. Moreover, just as this romantic delight may often provide an additional and distinct source of motivation to act in ways that please the person we are romantically attracted to, the delight of filial communion also provides its own unique kind of motivation to please God the Father in filial obedience.

An important aspect of the emotional experience of filial delight in the act of obedience is the believer's construal of God's emotional reaction toward his or her act of obedience. The pleasure of obeying God as Father is in significant part brought about by one's experience of God's joy or delight toward the virtuous obedience of his children. Such an experience of the Father's delight often requires

---

[55] *WJE* 10:641–642.
[56] Ibid.: 637.
[57] Ibid.: 640.
[58] Ibid.: 638.

an appropriate act of discernment and construal, on the part of the subject, of the emotional reaction of God toward his or her act of obedience. The appropriateness of the construal would, on the Christian tradition, be largely determined by the Christian Scripture, which has traditionally been taken to be God's self-revelation of his own character, purposes, and attitude toward his people.

The appropriate construal of God's emotional response to one's act of obedience together with the affective "coloring" brought about by the infused unitive drive, could generate a distinct kind of filial joy or delight in the believer's act of obedience to the Father. On the flip side, it could also generate a kind of filial grief in one's experience of filial disobedience and thus serve as a deterrence for disobedience. Such filial joy and grief, as suggested by Edwards, provide a distinct source of motivation that removes the psychological burden of God's commands and makes obedience seemingly "easy and pleasant."

One significant point regarding the account of Christian filial obedience just sketched is that for most people, the joy of filial obedience requires a significant degree of training and cultivation. In particular, it would require the development of a habit of discerning and construing God's emotional response to one's act of obedience and the learning of how to appropriately form such construals. Therefore, the cultivation of the virtue of Christian filial obedience would likely demand some process of disciplined learning, training, and practice before it could be fully formed. Nevertheless, this virtue is not one that can be forged through human discipline and self-effort alone. It is a virtue that taps into the Christological filial longing and attraction for the Fatherhood of God that is granted only in and through the infusion of grace.

Concluding Remark

The infusion of grace is foundational to the Christian life and the practice of Christological Sonship. Without this grace, one ultimately cannot enter into the life of Christological Sonship, even if he or she can meaningfully relate with God to some extent. This is not unlike the way in which one cannot really enter fully into married life without a sexual drive because of the kind of relationship marriage is. Similarly, one cannot truly enter into a relationship of filial communion with God as Father without the infusion of grace because of the kind of relationship Christological Sonship is. This filial relationship with God as Father is at the heart of Christian living and praxis, and every other aspect of the Christian life flows from and is constituted by the foundational reality of Sonship in Christ, including the moral and ethical formation of believers.[59]

[59] I am grateful for the editors' feedback on an earlier version of this chapter. Their insightful comments have made the present chapter vastly better than it otherwise would be. This chapter was made possible through the support of a grant from the Character Project at Wake Forest University and

Bibliography

Alston, William P. 1989 "The Indwelling of the Holy Spirit." In *Divine Nature and Human Language: Essays in Philosophical Theology*, 121–150. Ithaca, NY: Cornell University Press.

Augustine. *On the Trinity* The Nicene and Post-Nicene Fathers, First Series: Vol. 3, edited by Philip Schaff. Edinburgh: T&T Clark, 1997.

Caldwell, Robert W. III *Communion in the Spirit: The Holy Spirit as the Bond of Union in the Theology of Jonathan Edwards*. Milton Keynes: Paternoster, 2006.

Danaher, William J. *The Trinitarian Ethics of Jonathan Edwards*. Louisville, KY: Westminister John Knox Press, 2004.

Delattre, Roland. "Beauty and Theology: A Reappraisal of Jonathan Edwards." *Soundings: An Interdisciplinary Journal* 51, no. 1 (1968): 60–79.

Edwards, Jonathan. *Works of Jonathan Edwards.* Vol. 1, *Freedom of the Will*, edited by Paul Ramsey. New Haven, CT: Yale University Press, 1957.

———. *Works of Jonathan Edwards.* Vol. 2, *Religious Affections*, edited by John E. Smith. New Haven, CT: Yale University Press, 1959.

———. *Works of Jonathan Edwards.* Vol. 3, *Original Sin*, edited by Clyde A. Holbrook. New Haven, CT: Yale University Press, 1970.

———. "Sermon One: Love the Sum of All Virtue." In *Works of Jonathan Edwards,* vol. 8, *Ethical Writings*, 129–149, edited by Paul Ramsey. New Haven, CT: Yale University Press, 1989.

———. "Sermon Two: Love More Excellent Than Extraordinary Gifts of the Spirit." In *Works of Jonathan Edwards,* vol. 8, *Ethical Writings,* 149–174, edited by Paul Ramsey. New Haven, CT: Yale University Press, 1989.

———. "True Love to God." In *Works of Jonathan Edwards*, vol. 10, *Sermons and Discourses, 1720-1723*, 632–645, edited by Wilson H. Kimnach. New Haven, CT: Yale University Press, 1992.

———. *Works of Jonathan Edwards*. Vol. 13, *The Miscellanies: A-500*, edited by Harry S. Stout. New Haven, CT: Yale University Press, 1994.

———. "A Divine and Supernatural Light." In *Works of Jonathan Edwards, vol. 17, Sermons and Discourses, 1730-1733*, 405–427, edited by Mark Valeri. New Haven, CT: Yale University Press, 1999.

———. *Works of Jonathan Edwards.* Vol. 18, *The Miscellanies: 501-832,* edited by Ava Chamberlain. New Haven, CT: Yale University Press, 2000.

———. "The Excellency of Christ." In *Works of Jonathan Edwards,* vol. 19, *Sermons and Discourses, 1734-1738*, 560–595, edited by M. X. Lesser. New Haven, CT: Yale University Press, 2001.

———. "'Controversies' Notebook: The Nature of True Virtue." In *Works of Jonathan Edwards,* vol. 21, *Writings on the Trinity, Grace, and Faith*, edited by Sang Hyun Lee. New Haven, CT: Yale University Press, 2002.

---

the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

Edwards, Jonathan. "Discourse on the Trinity." In *Works of Jonathan Edwards,* vol. 21, *Writings on the Trinity, Grace, and Faith,* 109–145, edited by Sang Hyun Lee. New Haven, CT: Yale University Press, 2002.

———. "Treatise on Grace." In *Works of Jonathan Edwards,* vol. 21, *Writings on the Trinity, Grace, and Faith,* 149–198, edited by Sang Hyun Lee. New Haven, CT: Yale University Press, 2002.

Erdt, Terrence. *Jonathan Edwards, Art and the Sense of the Heart*. Amherst: University of Massachusetts Press, 1980.

Holmes, Stephen R. *God of Grace and God of Glory: An Account of the Theology of Jonathan Edwards*. Edinburgh: T&T Clark, 2000.

Lee, Sang Hyun. *The Philosophical Theology of Jonathan Edwards*. Princeton NJ: Princeton University Press, 1988.

McClymond, Michael J. *Encounters with God: An Approach to the Theology of Jonathan Edwards*. New York: Oxford University Press, 1998.

———. "Salvation as Divinization: Jonathan Edwards, Gregory of Palamas and the Theological Uses of Neo-Platonism." In *Jonathan Edwards, Philosophical Theologian*, 139–160, edited by Paul Helm and Oliver Crisp. Burlington, VT: Ashgate, 2003.

Miller, Perry. "Jonathan Edwards and the Sense of the Heart." *Harvard Theological Review* 41 (1948): 123–145.

Morimoto, Anri. *Jonathan Edwards and the Catholic Vision of Salvation*. University Park: Pennsylvania State University Press, 1995.

Ramsey, Paul. Editor's introduction to *The Works of Jonathan Edwards.*Vol. 8, *Ethical Writings*, 1–121. New Haven, CT: Yale University Press, 1989.

Smith, John E. Editor's introduction to *The Works of Jonathan Edwards.* Vol. 2, *Religious Affections*, 1–83. New Haven, CT: Yale University Press, 1959.

Walton, Brad. *Jonathan Edwards: Religious Affections and Puritan Analysis of True Piety, Spiritual Sensation and Heart Religion*. New York: Edwin Mellon Press, 2002.

Wynn, Mark. *Emotional Experience and Religious Understanding: Integrating Perception, Conception and Feeling*. New York: Cambridge University Press, 2005.

Yeo, Ray S. "Toward a Model of Indwelling: A Conversation with William Alston and Jonathan Edwards." *Journal of Analytic Theology* 2 (2014).

25 }

Necessity and Human Agency: Cultivating Character in the Reformed Christian Tradition

Elizabeth Agnew Cochran

Introduction

Ethicists working within any theistic tradition face the challenge of conceiving and articulating the roles of divine and human activity in forming a moral agent's character. This challenge is particularly pronounced for Reformed Protestant ethicists. Reformed theology has traditionally emphasized divine sovereignty over human affairs to a degree that critics of this tradition see as undermining human agency. Such an emphasis on divine sovereignty is evident in a Reformed understanding of providence and in the soteriological conviction that salvation is the work of God alone. This tradition's theological commitments give rise to a view of the moral life in which God plays a necessary and decisive role in fostering the development of good character, so much so that good character cannot be achieved apart from divine assistance.

This chapter explores possible contours of a conception of moral agency that is consistent with a broadly Reformed understanding of God's essential place in moral development but that simultaneously understands humans as fully accountable for the pursuit of a particular kind of moral character. In responding to the challenge of determining how to balance the roles of divine and human agency in the pursuit of virtue, I argue that Roman Stoic ethics provides a framework that merits further consideration. While certain dimensions of Stoic ethics have rightly been rejected both by historical and contemporary Christian theologians, a careful and judicious retrieval of select Stoic convictions provides a philosophical framework for developing an ethic that maintains an authentic account of human moral agency while retaining the Reformed commitment to a robust account of the work of divine providence. This ethic centers on a model of moral responsibility that is realized as human beings pursue two closely related character dispositions: first, an acceptance of

one's particular moral capacities as situated in relation to certain limitations of one's circumstances, and second, a love for strangers that emerges from a recognition of one's own place in the universe. I begin this chapter with a brief discussion of the Reformed theological claims that pose particular challenges for reflection on moral agency in humans. I then consider more fully the two dispositions central to an account of moral agency that works to accommodate these theological convictions.

Divine and Human Agency in Reformed Protestant Thought

Before laying out the character dispositions that form the foundation of a positive understanding of moral agency in a world closely overseen by divine providence, it will be helpful to recognize more fully the ways in which Reformed understandings of providence and human salvation elevate God's role in human moral formation. Sixteenth-century theologian John Calvin, a central figure of the Protestant Reformation whose ideas decisively shaped Reformed theology, adheres to an account of providence that emphasizes God's active and direct oversight of human affairs. Calvin characterizes God as the "disposer and ruler of all things."[1] He insists both that God radically transcends the created order and that God is deeply and immanently involved in world events. He characterizes God's will as "the supreme and primary cause of all things, because nothing happens without his [God's] order or permission."[2] Calvin contends that God does not withdraw from the world after creating it,[3] nor does God work providentially in a merely indirect way.[4] Instead, God remains active in the world after creating it and oversees and guides the world in a "specific" as well as a general sense. Calvin uses the term "special providence" to highlight God's pervasive oversight of the world.[5] God actively works "in the disposing and directing of every thing to its proper end by incomprehensible wisdom."[6] Calvin is confident that God's oversight is such that all events occur "by the ordination and command of God," and "particular events" reveal the workings of this providence.[7]

Not only does Calvin maintain that God is the direct and immediate cause of world events, but he more specifically understands salvation in terms that attribute a strong role to God's agency in enabling the development of moral

[1] Calvin ([1559] 2008, pp. 119–120), *Institutes of the Christian Religion*, 1.16.8.
[2] Ibid. (p. 120), *Institutes* 1.16.8
[3] Ibid. (p. 114), *Institutes* 1.16.1.
[4] Ibid. (pp. 116–119), *Institutes* 1.16.4–5; see also p. 113.
[5] Crisp (2010) explains the technical distinction between general and specific providence (p. 53).
[6] Calvin ([1559] 2008, p. 117), *Institutes* 1.16.4.
[7] Ibid. (p. 115), 1.16.3.

character. He adheres to a view of original sin as so strongly affecting humans' moral capacities that humans cannot do good apart from the personalized intervention of God's salvific grace. Prior to undergoing justification, which restores a person's relationship with God through the imputation of Christ's perfect righteousness to that individual,[8] a human being is a "sinner"[9] whose nature prevents her from being able to make any sort of "progress" toward the good.[10] Calvin attributes election to God's "will," God's "pleasure," and God's "secret counsel" rather than to human work or merit,[11] and election is a necessary and essential precursor to holiness.[12]

Eighteenth-century American Puritan theologian Jonathan Edwards stands in keeping with Calvin's suggestion that human capacities for holiness or goodness depend on God, underscoring the centrality of this conviction for historical Reformed theology. Edwards insists that genuine virtue can be practiced only by the elect, those whom God has justified.[13] He speaks of certain qualities, such as justice or a proclivity to love one's friends and family, as forms of a "natural good" that humans can pursue prior to undergoing justification.[14] But he nevertheless maintains that these qualities do not produce or even promote true virtue.[15] Indeed, God gives justified Christians a "spiritual sense," a new way of using their natural faculties, apart from which they cannot practice virtue.[16] For both Calvin and Edwards, then, God's will plays a central role in determining a person's moral capacities and moral character.

While Reformed theological convictions pose certain challenges for thinking through the contours of human moral agency, the Reformed moral vision does presume that humans are responsible for their behavior. Calvin and Edwards address questions of moral responsibility primarily through considering what it means to attribute free will to humans. Calvin, in keeping with his contemporary Martin Luther, draws on a distinction between the terms *necessitas* and *coactio* to affirm that the human will is free. Luther argues that people are subject to *necessitas immutabilitatis* that requires us to act consistently with our characters. But he simultaneously contends that we are "free" because we are not subject to external compulsion, or *necessitas coactionis.*[17] Calvin likewise suggests that *coactio*, compulsion or coercion by an external force, would limit freedom, but a necessity consisting in acting consistently

[8] Ibid. (pp. 475–477), *Institutes* 3.11.2; 3.11.4.
[9] Ibid. (p. 475), 3.11.2.
[10] Ibid. (pp. 617–618), 3.22.4.
[11] Ibid. (p. 618), 3.22.4; (pp. 618–619), 3.22.6.
[12] Ibid. (pp. 616–617), 3.22.3.
[13] See Cochran (2011b, 34–37).
[14] Edwards ([1765] 1989), *The Nature of True Virtue*, 568–569, 601–610.
[15] Ibid., 602.
[16] Edwards ([1746] 1959), *Religious Affections*, p. 206. See also McClymond (1995, 21–22).
[17] Martin Luther (1972, 64–65).

with one's nature does not problematically obstruct freedom.[18] He explains that God is free but is simultaneously subject to a sort of necessity in that he is necessarily good. He clarifies that this necessity does not mean that God is coerced by a "violent impulse." Instead, the fact that God "cannot do evil" is a consequence of God's "boundless goodness."[19] Likewise, humans are free "only to the extent" of being free "from compulsion";[20] otherwise, we are bound to the same necessity to which God is subject. Our internal nature necessarily shapes our actions, but Calvin contends that our actions themselves remain voluntary.

Edwards develops a similar argument concerning the will's freedom, framing the distinction between necessity and compulsion as one of "moral necessity" versus "natural necessity." Edwards explains that "natural necessity" involves the subjection of the will to natural limitations; for example, mathematical propositions are necessarily true, and humans are subject to the force of gravity. "Moral necessity," in contrast, refers to a necessary connection between "moral causes and effects." A moral cause would include a "moral habit or motive," and a moral effect would describe an act of the will.[21] Edwards argues that human beings are subject to moral necessity. Our desires and characters necessarily determine our actions so that we may find ourselves unable to act counter to our characters.[22] But this form of necessity differs from "natural necessity," which, like *coactio*, occurs when "extrinsic" forces prevent us from acting as we desire.[23] Edwards argues that our consciences help us recognize that actions subject to natural necessity are not praiseworthy or blameworthy.[24] But in thinking about moral necessity, we see an opposite correlation between necessity and praiseworthiness: actions are *more* praiseworthy or blameworthy when they are consistent with a person's dispositions than when they are not connected to them.[25] Like Calvin, Edwards argues for a moral necessity in God[26] that underscores the praiseworthiness of morally necessary actions. God is a "sovereign" being with "infinite *power*,"[27] but he is simultaneously subject to the moral necessity of being "holy and wise."[28] Both Calvin and Edwards attribute a certain kind of necessity to the human will and

[18] Calvin ([1559] 2008, p. 162), *Institutes* 2.2.7.
[19] Ibid. (p.181), 2.3.5.
[20] Ibid. (pp. 161–162), 2.6.
[21] Edwards ([1754] 1957), *Freedom of the Will*, 156–158.
[22] Ibid., 159–160.
[23] Ibid., 159.
[24] Ibid., 359.
[25] Ibid., 360.
[26] Ibid., 377.
[27] Ibid., 379.
[28] Ibid., 380–381. Edwards affirms here that God's will is not self-determining; instead, "God's perfect wisdom" determines his will.

connect this necessity to a person's moral character, internal dispositions that shape her actions and identity.

Necessity and Moral Responsibility: A Constructive Proposal for the Reformed Tradition

Ethicists working within the Reformed tradition are faced with a challenging legacy. Calvin's and Edwards's arguments for the connection between actions and dispositions are compelling, but their defense of this position subjects the human will to a certain kind of causal necessity: choices of the will are necessarily effects caused by a person's dispositions. Our nature determines who we are and dictates the sorts of actions we are capable of pursuing. This position is further complicated by Calvin's and Edwards's soteriological commitments that identify God as the agent whose will solely determines the presence or absence of moral capacities in a given human person. If God implants these capacities in a person and these capacities necessarily give rise to particular character traits, then it would seem that human agency is severely restricted.

Yet even as Calvin's and Edwards's understandings of the limitations of human capacities for goodness are central to this tradition, it remains simultaneously a central Reformed theological tenet that God is good and that the divine essence is intrinsically loving. Various historical thinkers outside the Reformed tradition, including the Stoic Epictetus and the Renaissance humanist Erasmus, have argued that such a position requires that humans exercise a certain kind of freedom for their actions: otherwise God might appear cruel and arbitrary in making certain judgments,[29] or it might appear that God is the author of evil.[30] Both of these claims about God would be theologically problematic, and one way of guarding against them is to develop a cogent account of precisely how people remain morally accountable for their actions even as God plays a strong role in determining their moral capacities.

Although Reformed theologians have historically recognized the importance of attributing moral accountability to people,[31] certain of their positions

[29] Epictetus's commitment to humanity's moral freedom is crucial to preserving his conception of God as caring for human beings: "For if God had so constructed that part of His own being which He has taken from Himself and bestowed upon us, that it could be subjected to hindrance or constraint either from himself or from some other, He were no longer God, nor would He be caring for us as He ought" (1998, 1:117–119), *Discourses* I.17.

[30] Erasmus's discussion of God's hardening of Pharaoh's heart illustrates this concern. Erasmus argues that we must interpret this "hardening" as an act of God's foreknowledge of Pharaoh's obstinance rather than an act through which God imposes a disposition on Pharaoh against Pharaoh's will. Pharaoh, he explains, "was not forced by the will of God to be obstinately wicked" (1969, 65–67), for this would be incompatible with God's justice and goodness.

[31] Edwards ([1754] 1957) is particularly clear on this point, *Freedom*, 269, 368–369.

risk falling prey to the problems that can be associated with failing to articulate a robust account of moral agency in humans. For example, Oliver Crisp contends that Edwards risks presenting God as the author of evil.[32] Jennifer Herdt, in turn, argues that Martin Luther is unable to develop a satisfying account of moral agency in humans. She explains that not only does Luther insist upon a moment of passivity before God as the starting point for virtue,[33] but he also remains deeply suspicious of the agency of even justified Christians. Genuine human agency is always, for Luther, a problematic assertion of self-love. Even after a Christian has been justified, human agency is intrinsically "prideful," and its exercise is, according to Herdt's reading of Luther, "inherently an attempt to displace divine agency."[34] For Luther, Herdt explains, the Christian life requires "displacing" our own agency and embracing a pure "passivity" that allows God to be the exclusive agent who enacts our good actions.[35] Herdt's concerns about Luther implicitly challenge ethicists working in the Reformed tradition to consider carefully how one might go about conceiving a genuinely participative understanding of human moral agency without entirely rejecting a traditional Reformed understanding of providence and the idea that humans depend radically on God. The remainder of this chapter lays out a vision of moral agency in humans that can function within a Reformed theological framework.

The account of human agency developed here is centered on two interrelated character dispositions that a Reformed theological context implicitly invites and encourages a moral agent to pursue. My discussion of each disposition is heavily informed by the Roman Stoics. Many of the Stoics' metaphysical beliefs are deeply at odds with Reformed theology, and the Stoics' understanding of providence differs from Christianity's commitment to God's transcendence and the Christian belief that divine providence works in personalized and specific ways. However, the Roman Stoics share noteworthy moral commitments with Christianity, including a belief that divine providence actively works in the world directing it toward its good and that virtue is embodied in recognizing and assenting to God's providential oversight of the universe.[36] Stoic ethics has historically been attractive to a number of Reformed theologians,[37] and

[32] Crisp (2005, 13, 18, 21, 54–78).

[33] Herdt (2008, 174–175).

[34] Ibid., 188.

[35] Ibid.

[36] For further consideration of the continuity and discontinuity between the Roman Stoic account of virtuous assent and the understanding of faith at work in Martin Luther, John Calvin, and Jonathan Edwards, see Cochran (2014).

[37] Several ordained Presbyterian pastors affiliated with the Scottish Enlightenment, such as Francis Hutcheson, George Turnbull, and Hugh Blair, carefully and explicitly engaged the Roman Stoics. For more discussion of these thinkers and their interplay of Stoic and Christian moral commitments, see Haakonsen (1996); and Sher (1990). For more on possible intellectual similarities between Stoic positions and specific Reformed theologians, see Cochran (2011a); and Crouter (1974).

these theologians' historical interest in the Stoics points toward the possibility that Stoic thought might have the potential to illuminate contemporary conversations about the challenges of articulating the relationship between God's divine sovereignty and humanity's moral accountability. The Roman Stoics develop a thoroughgoing account of moral responsibility, and a judicious retrieval of this account provides a starting point for positing a conception of human moral agency that is consistent with a belief that divine providence subjects human beings to certain sorts of necessity.

Acceptance of Limitations

The first disposition through which humans actively cultivate a moral character while remaining faithful to Reformed understandings of providence and necessity is a deliberative recognition of their own moral limitations. The idea that human beings are necessarily subject to certain limitations or constraints is consistent with the broader aims of the contemporary philosophical and theological movement known as virtue ethics. This movement has focused on retrieving ancient moral traditions in order to challenge modern understandings of ethics that tend toward a view of the human person as an autonomous will that can be detached from historical and material particulars.[38] Like contemporary virtue ethics, the Reformed theological tradition runs counter to an autonomy-focused understanding of personhood by presuming that humans are necessarily limited in their moral capacities. Indeed, recognizing one's own inability to achieve salvation and acknowledging one's complete dependence upon God is often presented as a crucial disposition in narratives of conversion to Christianity. In addition to Luther's account of conversion discussed briefly earlier, Edwards's narrative *The Life of David Brainerd* describes an event crucial to Brainerd's conversion as one in which Brainerd becomes captivated by a vision of God's perfections. This vision enables Brainerd to recognize the disparity between the utter beauty of divine goodness and the lack of merit in human works.[39] Edwards's association of Brainerd's awareness of human moral limitations with Brainerd's conversion suggests that such awareness is deeply important to one's Christian journey.

When considered solely in the context of conversion, it might seem that recognizing our moral limitations necessitates a complete renunciation of one's own moral agency and replacement of this agency with the activity of

[38] The coupling of this retrieval of ancient traditions with concerns about problematic tendencies of Kantian and utilitarian ethics is illustrated in several major texts associated with the beginnings of contemporary virtue ethics. For philosophical sources, see Anscombe (1958) and MacIntyre (2007). For theological sources, see Hauerwas (1991).

[39] Edwards ([1749] 1957), *Life of David Brainerd*, 132–134.

God's will. In this very narrow context, it would be appropriate for a Christian to think of her recognition of dependence on God for salvation as a recognition that she is unable, through her own power, to overcome her sin and accomplish a reconciliation with God. But the Christian life extends beyond an initial moment of conversion. It is plausible to draw a distinction between the agency appropriate to conversion and that describing the activity of a justified Christian who opens herself to allowing Christ to transform her character. Living out the Christian faith requires cultivating a different sort of recognition of dependence on God, one that is embodied in a search to fulfill a particular vocation appropriate to one's own place in the universe God created and providentially oversees. A Christian exercises authentic moral agency through developing a proper understanding of herself in relation to God and seeking to live a life appropriate to her place in the universe.

The self-understanding central to this account of moral agency has a significant cognitive component. It is rooted in self-examination, in intellectual reflection on oneself as part of the created order. As creatures existing within the world, we are not autonomous and detached wills capable of controlling the things around us. Moral growth requires the pursuit of dispositions that equip us both to accept our lack of control over certain dimensions of life and to pursue the good in relation to those dimensions of life for which we hold responsibility.

The Roman Stoic Epictetus provides a helpful starting point for reflecting on how discernment about our limitations is essential to the development of a good moral character. The Stoics generally affirm that the universe is directed by a divine nature or natures who orchestrate and promote the good of humanity.[40] The Roman Stoics in particular emphasize the ways in which the divine providence at work in the universe is rational;[41] Epictetus characterizes God as intimately present in the world and overseeing all events.[42] Yet, while he implies that people can trust that God's activity in the world is surely directed toward humanity's good, Epictetus also affirms that individual human beings cannot control many dimensions of life. He presents reflection on which matters in life are under an agent's control as a crucial step in the achievement of virtue.

Epictetus recognizes that many dimensions of life are outside our control, but he affirms that humans do have the power to give or withhold assent to particular impressions that confront us. This power lies at the heart of moral agency as the Stoics understand it. Zeno, the founder of the first Stoic school,

[40] According to Cicero, the Stoics believe that the gods are intelligent beings who "set up and ordered" the laws of the natural world and that the gods "love us" in that they are "beneficent and friendly to mankind" (1:260).

[41] Annas (1993, 161).

[42] Epictetus (1998, 1:99–101), *Discourses* I.14.

argues that individuals have the ability to decide which sensory impressions merit mental assent.[43] In keeping with earlier Stoics, Epictetus reiterates that assent alone is the act that humans can choose to perform or not: the gods have given humans "the power to make correct use of external impressions," but leave all other things outside human "control."[44] We are morally responsible for recognizing the good and evil in the perceptions we receive and for choosing to assent to the good. For Epictetus, "the assents" through which our mind accepts particular impressions "are subject to man's will, and fall under his control."[45] We lack control over the faculties we possess or the sensory impressions that confront us, but we do have some measure of freedom in our response, in our "use" of the impressions we see, and in our decision to give them assent. Our ability to assent to circumstances lies at the heart of a Stoic account of moral responsibility. As Steven K. Strange puts it, "According to the Stoic view, it is primarily assent that is 'in our power' or 'up to us' (*eph' he min*), and it is only because this lies in our control that anything at all does. So if we are responsible for any of our actions, for any of our desires, or for any of our passions or emotions, it is because these all depend upon our capacity to give or withhold our assent in particular cases. Similarly, if we are to bear any responsibility for our own moral character, this too will be due to our capacity for assent."[46]

Epictetus's affirmation that assent lies within our power ultimately leads toward a conception of moral character as more fully under human control than historical Reformed positions would allow. For the Roman Stoics, virtue is embodied in assent to God's providence, so the affirmation that assent lies within our power is tantamount to an affirmation that we are capable of becoming virtuous through our own natural abilities. This conclusion is clearly at odds with Reformed theology; the Reformed tradition would hold that not even virtue is exclusively under our control but depends on God's will. Yet despite this important difference, Epictetus's advocacy of the practice of discerning and focusing on things that the gods have chosen to be "up to us" is nevertheless instructive because it gives rise to an understanding of moral agency that lies partly in recognizing oneself as subject to some measure of necessity caused by divine providence. Moreover, the virtue that Epictetus presents as being under human control consists in giving one's assent to the conviction that circumstances outside one's control are being directed by providence toward an end that is good for the world as a whole.[47] This

[43] Diogenes Laertius (1925, 2:157).

[44] Epictetus (1998, 1:9), *Discourses* I.1.

[45] Ibid. (2:449–453), fragment #9 in *Discourses*,

[46] Strange (2004, 35). Similarly, Sorabji (2000) argues that the Stoics see "voluntary assent" as a part of judgment, and explains: "The voluntariness is based on the idea that one is free to question appearances and withhold assent from them. If you do not bother to do so, that is your own fault" (45–46). Nussbaum (2001) likewise affirms that the Stoics "thought that assent was always a voluntary act, and that we always had it in our power to assent or refuse assent to any appearance" (38).

[47] Epictetus presents a cognitive conviction that "there is a God, and that his providence directs the whole" as a starting point for virtue (1998, 1:301), *Discourses* II.14. But he simultaneously makes clear

understanding of virtue resonates with the faith that Protestants see as definitive of a Christian's status before God. Like Roman Stoic virtue, Christian faith has a cognitive dimension but more fully consists in an embodied disposition of trust in God.[48]

A Reformed conception of human moral agency requires reconceiving our understanding of what it means to be a moral agent even more radically than Epictetus's position does. Moral agency does not ultimately lie in pursuing a virtue that is accessible to all persons through the construction of our nature, as virtue is for the Stoics.[49] Instead, it lies in recognizing that even virtue is not something our moral capacities can achieve apart from the assistance of God whose being is equivalent to virtue itself. Christians find their place as authentic moral agents by pursuing and embodying a disposition that acknowledges and accepts their own limitations, and particularly their moral limitations. A Reformed understanding of conscience indicates that we have the capacity to pursue this recognition through our own natural resources even prior to conversion, underscoring the plausibility of associating an awareness of our limitations with an exercise of agency that is properly "ours."[50] An acceptance of the notion that even moral matters lie outside our control lies at the heart of a Reformed understanding of moral agency. This conception of agency preserves the theological convictions that God providentially directs the world and that God is the author of human salvation. At the same time, elevating this disposition provides a starting point for ensuring that humans retain moral responsibility for working to pursue a good character even as this pursuit must be undertaken with humility.

---

that virtue requires giving one's assent to this proposition, and assent is tied to the faculty of the *prohairesis*, an Aristotelian term that the Roman Stoics employ to refer to a moral faculty that draws will and reason or intellect closely together. For interpretations of the Stoic *prohairesis* as a moral faculty that is not merely equivalent to the will, see Rist (1977, 231–232); Sorabji (2000, 215).

[48] Faith has roots in cognitive belief regarding God's character but is more fully embodied in a disposition of trust in God. See, for example, Gustafson (1977, 64, 45); Adams (1987, 20–23).

[49] The Stoics are committed to the equality of persons, and this equality is realized partly in the idea that all human beings possess an equal capacity for virtue. Seneca explains that each of us possesses reason, which gives us the ability to pursue virtue: "So powerful is virtue, and so gracious is Providence, that every man has a light set up within him for a guide." Virtue is therefore "open to all, as well to servants and exiles as to princes" (1984, 19, 22).

[50] To be sure, Reformed theology would suggest that God's justifying grace enables converted Christians to exercise this awareness of our dependence on God and our incapacity for virtue apart from God more fully than the unregenerate can. But conscience makes us aware of our moral limitations even prior to conversion and plays an important role in fostering dispositions appropriate to it. Edwards's account of repentance is helpful for considering the differences between these dispositions. He at some points speaks of repentance as an act an individual performs prior to her conversion, and at other points characterizes it as a virtuous disposition that is a counterpart of virtuous love accessible to the elect; as love for God grows, a regret and hatred for our remaining sin also increases. For an extended discussion of Edwards on repentance, see Cochran (2011b, 92–123).

The Stoics and the Reformers adhere to very different understandings of which things in life are "up to us." But Epictetus's alignment of moral responsibility exclusively with matters under our control provides a helpful framework for thinking of moral agency as authentic even when this agency is limited in scope. The Stoic aspirant to virtue is dissuaded from dwelling on dimensions of life that she cannot control. The Christian, more radically, does well to accept her lack of control, not only over particular historical and social circumstances that shape her life, but also over her relationship with God, for Christianity affirms that God's desire and capacity to love and forgive exceed human abilities to earn such love. Yet it should be stressed that for both Roman Stoicism and Christianity, the decision to cede control to divine providence is not an act of passive resignation to an arbitrary fate. Instead, this giving up of control is an expression of trust that the divine nature is benevolently ordering the world toward its good.

Love for Universal Being

The practices of self-examination that guide an individual in recognizing her limitations in achieving the good may appear to promote a view of the moral life that is overly individualistic and focused inward. Gilbert Meilaender expresses concern that virtue ethics as a discipline encourages moral agents to emphasize personal development in a manner that risks promoting an excessive preoccupation with oneself.[51] But while a concern with one's own moral limitations and discerning one's capacities for virtue does require an individual to turn inward, the Reformed Christian fully realizes her moral agency only through achieving a self-understanding that recognizes that she, as a human being, is deeply connected to all other beings in the universe. This self-understanding gives rise to a second disposition through which moral agency is realized: a universal love for God and for all of God's created order.

Multiple works in contemporary ethics call for a view of personhood that recognizes that all persons are formed by relationships with others and depend on others at certain stages of life. This shared experience of dependence indicates that humans are interrelated rather than isolated beings.[52] A Reformed conception of the necessary interrelation of human beings coincides with this scholarship in suggesting a view of the person as vulnerable and malleable; but it roots this view in particular theological claims about God and God's relationship to the created order rather than in reflection on human experience.

---

[51] Meilaender (1984, 13).

[52] Recent work that emphasizes dependence and vulnerability as central to human being includes Kittay (1999); MacIntyre (1999); Wendell (1996).

God is good and imparts divine goodness to the world, and to human beings specifically, in the act of creation. Original sin deeply limits humans' capacities to pursue moral goodness apart from God's gracious intervention. Nevertheless, all human beings share a nature that reflects the God who created them, and our shared nature and shared status as God's creatures give us a kinship with all other persons. This kinship challenges us toward a disposition of concern not simply for those in our immediate circles but for humanity as a whole.

The idea that people have moral obligations toward universal humanity is a moral commitment associated with Stoic thought. As Martha Nussbaum explains, the Stoics uphold a radical sense that the *cosmos* as a whole should be an individual's primary moral attachment, and a person's particular moral communities should be viewed as secondary.[53] Epictetus's account of virtue provides a framework that exemplifies how the moral disposition discussed in the preceding section of this chapter—a disposition that involves giving one's assent to circumstances outside one's control—is integrally linked to a disposition of universal concern for others. As noted above, virtuous assent is a disposition that emerges from an intellectual recognition that providence is good. As a moral agent recognizes the goodness of providence, she consents to her place in the universe and comes to understand the ways in which all parts of the universe are related. Epictetus uses the term *sympatheia* to describe the intimate ways in which the spiritual and material world reflect the character of divine providence.[54] The spiritual and material worlds are closely bound together, he contends, so that "what is on earth feels the influence of that which is in heaven." We see this influence in such patterns as the growth of flowers, the waxing and waning of the moon, and the ripening of fruit, all of which occur "regularly, as if from God's command."[55] Epictetus explains that just as "plants and our own bodies" are "closely" and "intimately" connected to the universe, this connection is even more present in our souls. In keeping with other Stoics who speak of God and the world as coextensive, Epictetus affirms that our souls, indeed, are so closely related to God as to be "parts and portions of His being," so that God can feel their "every motion" as something happening within himself.[56]

Although Epictetus's position indicates a pantheism that stands at odds with orthodox Christian theology, his account of an appropriate response to recognizing our close relation to other parts of the created order points toward a moral disposition fitting for the Reformed tradition as well: a spirit of concern for the other emerging from an understanding of persons as more than isolated individuals. Crucial to Stoic moral formation is the process of

---

[53] Nussbaum (1996, 345).
[54] Epictetus (1998, 1: 41–43), *Discourses,* I.6.
[55] Ibid. (1:99–101), I.14.
[56] Ibid., 1:101), I.14.

oikeiosis, through which aspirants to virtue grow in their recognition that all humans are, in a sense, members of the same household.[57] This understanding of commonality with other creatures leads the Stoics toward the position that humans have a moral obligation to exercise an impartial and universal concern for all other persons, a feeling that Richard Sorabji describes as a feeling of "good will (*eunoia*), defined as willing good to another for the other's sake."[58]

While the Stoics do not speak of *eunoia* precisely as love (although Sorabji argues that love is a subspecies of Stoic *eunoia*),[59] Reformed Christianity challenges people to love in a universal sense, and the Stoic account of willing the good of the other provides an entry point for understanding the nature of this love. Edwards's theology gives rise to an ethic that demands universal love and roots this ethic in the belief that God's creation of the world and of human beings within the world binds humans all together. As George Marsden puts it, "The key to Edwards's thought is that everything is related because everything is related to God."[60] Edwards affirms that human beings live out the purposes for which we were created when we practice a virtue that both reflects and participates in God's own goodness, which is expressed and embodied in a love for God and for all of God's creatures.[61] In spelling out the character of this virtue, Edwards is clear that genuinely virtuous love or "benevolence" must be directed toward "Being in general," both God and all of creation, rather than simply toward one's own family or friends or other persons whom one directly encounters[62] (thus feelings of pity or gratitude toward specific persons are not, strictly speaking, virtuous).[63] Edwards argues that virtue consists most essentially in "love to God," both because God is "infinitely the greatest and best of beings" and because God is the origin of all goodness and existence.[64] This love for God, in turn, necessarily gives rise to an awareness of one's interrelation with all other beings. A virtuous agent exercises "consent" toward God and toward the created world, and this consent establishes a "union" of a virtuous person's "heart" with the rest of creation.[65]

Edwards's account of the union of a moral agent's heart with the created order demonstrates the priority of love within a Reformed ethic. The character of this love partly coincides with Stoic *eunoia*. The Stoic commitment to willing the good of another person for that person's sake indicates that virtuous love must be unselfish and that this love must recognize and celebrate the intrinsic value of each individual human being. But Christian moral

---

[57] For more on *oikeiosis*, see Sorabji (2000, 174).
[58] Ibid.
[59] Ibid.
[60] Marsden (2003, 460).
[61] Cochran (2011b, 154–155); Edwards ([1765] 1989), *Nature*, 439, 503.
[62] Edwards ([1765] 1989), *Nature*, 540.
[63] Ibid., 601–610.
[64] Ibid., 550–555.
[65] Ibid., 540.

thought also affirms, as we see in Edwards's account of benevolence, that love for God and love for one's neighbor are so closely related as to be dimensions of the same singular disposition. Universal love must capture something of this sense that love for another recognizes the other's status as a creature of God, so that in loving another person, we are living out a theological conviction that God creates and values particular individuals. One line of thought in Eric Gregory's *Politics and the Order of Love* illustrates this idea. In this work, Gregory explores Augustine's consideration of the question of whether a person should love her friend for the friend's sake or for the sake of God and ultimately defends the idea that Christians should love other persons for God's sake.[66] A Christian ethic demands a love for the other specifically in that individual's relationship to God. Stoicism is open to such an ethic insofar as concern for others emerges from a recognition of shared kinship in the world community, but Christian convictions press us to make a central focus on God explicit in thinking through precisely how this love should be understood. Christian love is embodied in a concern for the other not simply as sharing a kind of creaturely kinship with us through our shared humanity, but more specifically for the other as a creature whose existence, like ours, is determined through a relationship with God.

While Gregory's account of love for other persons as creatures of God speaks effectively to the nature of Christian love, Reformed soteriological convictions require an account of an individual's formation in this virtuous love that differs from Gregory's view of this formation. Gregory suggests that personal relationships provide a starting point for cultivating universal love,[67] a position consistent with Diana Fritz Cates's Aristotelian-Thomist account of how humans are formed in the virtue of compassion.[68] Reformed commitments, in contrast, require thinking of universal love for God and neighbor as a starting point for practicing specific and personal loves in a moral way, and Stoic ethics provides a philosophical framework that accommodates this difference well. Christian love, like Stoic virtue, begins with a recognition that humans all share a relation to God that involves partaking in God's nature, and this recognition generates a love for others that cannot be detached from core Christian beliefs regarding God's character. For Edwards, love for friends and family cannot lead us toward the exercise of virtuous benevolence. Benevolence is of a different order than these other loves and must be present in order for someone's exercise of these loves to be considered moral.[69] The Stoics

[66] Gregory (2008, 333–343). Gregory notes that some scholars have read this idea as instrumentalizing other persons but calls for an interpretation of Augustine's terms *uti* and *frui* that guards against this interpretation (337–338).

[67] Gregory (2008, 294–296).

[68] Cates (1996).

[69] For more discussion, see Cochran (2011b, 124–166).

likewise conceive universal love as a necessary foundation for the proper exercise of specific loves. Epictetus explains that virtue is necessary in order for us to be a good friend, parent, or sibling: "If, therefore, I am where my moral purpose is, then, and then only, will I be the friend and son and the father that I should be."[70]

The idea that human beings are morally obligated to pursue a disposition of universal love for other persons may seem abstract and impractical. It is important, however, to remember that in the Reformed tradition, this disposition stands alongside a conviction that providence so thoroughly oversees human affairs that God is concerned for and invested in individual experiences and encounters. While an ideal of universal love urges Christian moral agents to actively seek out settings in which compassion for others is particularly required, a Reformed account of providence simultaneously suggests that God provides opportunities for each person to offer universal love toward the specific others whom he or she encounters in daily life. Universal love is an abstract ideal, but a disposition of universal love can be pursued through concrete exchanges and relationships. In the particular settings of their day-to-day lives, moral agents are given opportunities to love others and to will their good as creatures of God whose lives possess intrinsic value.

Conclusion

The character dispositions outlined here are rooted in Stoic moral thought and demonstrate how this tradition offers resources for envisioning the contours of human moral agency within a Reformed theological framework. A belief in God's sovereignty and a conviction that God alone is the author of salvation require that a Christian conceive moral goodness as something that humans cannot pursue or cultivate exclusively through their own efforts. These convictions place certain necessary restrictions on human moral agency, but I have argued here that such agency can be exercised authentically through cultivating moral dispositions that embody an acceptance of our limitations and a love for all human beings. A disposition of accepting one's limitations opens an individual to the moral possibilities suited to her own particular circumstances and character and ensures that she retains a focus on matters for which she stands responsible. A proper discernment of one's own limitations and capacities, in turn, fosters a self-understanding that gives rise to a sense of oneself as a being in the universe. This self-understanding is most fully and completely embodied in universal love. These dispositions together shape a moral character that provides a starting point for understanding how

[70] Epictetus (1998, 1:389), 2.22.

ethicists working in the Christian tradition can preserve moral accountability in humans even while conceiving the pursuit of the good as in some ways outside human control.[71]

References

Adams, Robert Merrihew. 1987. *The Virtue of Faith and Other Essays in Philosophical Theology*. New York: Oxford University Press.

Annas, Julia. 1993. *The Morality of Happiness*. New York: Oxford University Press.

Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy* 33 (124): 1–16.

Calvin, John. (1559) 2008. *Institutes of the Christian Religion*. Translated by Henry Beveridge. Peabody, MA: Hendrickson Publishers.

Cates, Diana Fritz. 1996. *Choosing to Feel: Virtue, Friendship, and Compassion for Friends*. Notre Dame, IN: University of Notre Dame Press.

Cicero.1987. *On Divination.* Vol. 1, *The Hellenistic Philosophers*, edited by A. A. Long and D. N. Sedley, I:82–83. New York: Cambridge University Press..

Cochran, Elizabeth Agnew. 2011a. "Consent, Conversion, and Moral Formation: Stoic Elements in Jonathan Edwards's Ethics." *Journal of Religious Ethics* 39 (4): 623–650.

Cochran, Elizabeth Agnew. 2011b. *Receptive Human Virtues: A New Reading of Jonathan Edwards's Ethics*. University Park: Pennsylvania State University Press.

Cochran, Elizabeth Agnew. 2014. "Faith, Love, and Stoic Assent: Reconsidering Virtue in the Reformed Tradition." *Journal of Moral Theology* 3 (1): 199–227.

Crisp, Oliver. 2005. *Jonathan Edwards and the Metaphysics of Sin.* London: Ashgate.

Crisp, Oliver. 2010. *Retrieving Doctrine: Essays in Reformed Theology*. Downers Grove, IL: Intervarsity Press.

Crouter, Richard E. 1974. "H. Richard Niebuhr and Stoicism." *Journal of Religious Ethics* 2 (2): 129–146.

Diogenes Laertius. 1925. *Lives of Eminent Philosophers*. 2 vols. Translated by R. D. Hicks. Cambridge, MA: Harvard University Press.

Edwards, Jonathan. (1754) 1957. *Freedom of the Will.* Vol. 1, *The Works of Jonathan Edwards,* edited by Paul Ramsey. New Haven, CT: Yale University Press.

Edwards, Jonathan. (1749) 1984. *The Life of David Brainerd*. Vol. 7, *The Works of Jonathan Edwards*, edited by Norman Pettit. New Haven, CT: Yale University Press.

Edwards, Jonathan. (1765) 1989. "The Nature of True Virtue." In *Ethical Writings,* vol. 8, *The Works of Jonathan Edwards*, edited by Paul Ramsey. New Haven, CT: Yale University Press.

[71] This publication was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. I am also grateful to Christian Miller and Angela Knobel for their comments on an earlier version of this chapter and to all those who attended my presentation at the 2013 conference "Conclusions about Character: Results from Psychology, Philosophy, and Theology," whose feedback helped me in developing these arguments.

Edwards, Jonathan. (1746) 1959. *Religious Affections*. Vol 2, *The Works of Jonathan Edwards*, edited by John E. Smith. New Haven, CT: Yale University Press.

Epictetus. 1998. *The Discourses as Reported by Arrian*. 2 vols. Translated by W. A. Oldfather. London: Harvard University Press.

Erasmus, 1969. *On Free Choice of the Will*. In *Luther and Erasmus: Free Will and Salvation*, edited by E. Gordon Rupp and Philip S. Watson. Louisville, KY: Westminster John Knox Press.

Gregory, Eric. 2008. *Politics and the Order of Love: Toward an Augustinian Ethic of Democratic Citizenship*. Chicago: University of Chicago Press.

Gustafson, James. 1977. *Can Ethics Be Christian?* Chicago: University of Chicago Press.

Haakonsen, Knud. 1996. *Natural Law and Moral Philosophy: From Grotius to the Scottish Enlightenment.* Cambridge: Cambridge University Press.

Hauerwas, Stanley. 1991. *A Community of Character: Toward a Constructive Christian Social Ethic*. Notre Dame, IN: University of Notre Dame Press.

Herdt, Jennifer A. 2008. *Putting on Virtue: The Legacy of the Splendid Vices.* Chicago: University of Chicago Press.

Kittay, Eva Feder. 1999. *Love's Labor: Essays on Women, Equality, and Dependency*. New York: Routledge.

Luther, Martin. 1972. *On the Bondage of the Will*. In *Luther's Works,* vol. 33, *Career of the Reformer III*. Minneapolis, MN: Fortress Press.

MacIntyre, Alasdair. 2007. *After Virtue: A Study in Moral Theory*. 3rd ed. Notre Dame, IN: University of Notre Dame Press.

MacIntyre, Alasdair. 1999. *Dependent Rational Animals: Why Human Beings Need the Virtue.* Peru, IL: Open Court Press.

Marsden, George. 2003. *Jonathan Edwards: A Life*. New Haven, CT: Yale University Press.

McClymond, Michael J. 1995. "Sinners in the Hands of a Virtuous God: Ethics and Divinity in Jonathan Edwards's *End of Creation*." *Zetischrift fur neuere Theologiegeschichte* 2: 1–22.

Meilaender, Gilbert C. 1984. *The Theory and Practice of Virtue*. Notre Dame, IN: University of Notre Dame Press.

Nussbaum, Martha C. 1996. *The Therapy of Desire*. Princeton, NJ: Princeton University Press.

Nussbaum, Martha C. 2001. *Upheavals of Thought: The Intelligence of Emotions*. New York: Cambridge University Press.

Rist, J. M. 1977. *Stoic Philosophy*. Cambridge: Cambridge University Press.

Seneca. 1984. *Of a Happy Life.*In *The Wisdom of the Stoics: Selections from Seneca, Epictetus, and Marcus Aurelius*, edited by Frances Hazlitt and Henry Hazlitt. Lanham, MD: University Press of America.

Sher, Richard B. 1990. *Church and University in the Scottish Enlightenment: The Moderate Literati of Edinburgh*. Edinburgh: Edinburgh University Press.

Sorabji, Richard. 2000. *Emotion and Peace of Mind: From Stoic Agitation to Christian Temptation*. Oxford: Oxford University Press,

Strange, Steven K. 2004. "The Stoics on the Voluntariness of the Passions." In *Stoicism: Traditions and Transformations*, ed. Stephen K. Strange and Jack Zupko, 32–51. New York: Cambridge University Press.

Wendell, Susan. 1996. *The Rejected Body: Feminist Philosophical Reflections on Disability.* New York: Routledge.

26 }

Liturgy and the Moral Life

Terence D. Cuneo

If you attend the performance of most Christian liturgies, you will find the assembled reading and listening to the scriptural depiction of various events that compose the core Christian narrative. If you attend the performance of a liturgy of the Eastern Christian churches, however, you will discover that the assembled do not simply read or listen to the scriptural presentation of the events that compose the core Christian narrative; they also reenact many of these events by imitating them. During the services of Holy Week, for example, the assembled will read and listen to what the Gospels say about the events surrounding Jesus's death. Then, at various points, they reenact some of these events, such as Jesus's washing of the disciples' feet and his burial. At these points, the readers and listeners become performers.

The question that I would like to pursue in this chapter is why this is so. Why would the scripts of the ancient liturgies direct those assembled for the liturgy, not merely to read and listen to the scriptural presentation of various events that compose the core narrative, but also to perform actions that imitate them? In raising this question, I am not interested in identifying causal or historical explanations of why the activity of reenactment figures so prominently in the ancient liturgies, as interesting as such explanations might be. Instead, I am asking what the activity of liturgical reenactment is *for*, what its contribution to the moral and religious life is supposed to be.

The answer that I am going to offer assumes that there is no single contribution that liturgical reenactment is supposed to make to the moral and religious life; there are probably multiple ways in which reenactment is supposed to contribute to the realization of various ethical and religious ideals. Still, I suggest that we can helpfully speak of the dominant ends or goals of liturgical reenactment, and that among these dominant ends is that of contributing to the construction of a narrative identity of the participants.

By an agent's narrative identity, I have in mind a sequence of events which has that agent as a subject, to which he or she might refer if he or she were accurately to tell a story of his or her life. Suppose, for example, I were to ask you a series of questions, such as who are you, what made you who you are, and what do you value most in your life? You might answer by telling me that you are the one whose love of music led to years studying the cello and whose dedication to the instrument required sacrificing a promising dancing career. Your having such properties as *being a lover of music* would be part of your narrative identity, since it would form part of a storylike sequence of events, which has you as a subject, to which you might refer were you to tell a story of your life. (I do not assume that there must be only one such story you could offer in response to my questions, or that it would have to fall into some specific genre or have any specific content. Such a story simply needs to be responsive to the types of questions I just mentioned.)

In focusing my discussion on the topic of narrative identity, I realize that I am joining a much larger conversation about the role of narrative in the moral life. The contribution I hope to make to this discussion is not, however, to advance another descriptive claim about the ways in which certain activities are likely to contribute to the development of one's narrative identity. So, for example, I am not going to develop a variant of the thesis that Plato railed against—namely, that imaginatively identifying with a character in some work of art, such as a drama, increases the likelihood that one becomes like that character, thereby undermining self-control. Instead, the view I defend is thoroughly normative. I am interested in the normative commitments that agents undertake when engaging in liturgical reenactment and how these commitments might contribute to the construction of a narrative identity of a particular sort.

My topic in this chapter is a large one, so I should call attention to several respects in which my discussion will be selective. Although the activity of liturgical reenactment is prominent in the ancient liturgies, I am going to focus on, not these liturgies in general, but only those of the Eastern Christian churches, as these strike me as a particularly rich resource to explore given our topic. Moreover, my eye will be on, not the totality of the Eastern liturgies, but on the Lenten liturgies in particular, as it is in these services, which are deeply penitential, that the activity of liturgical reenactment figures most prominently. For ease of reference, I will often refer to "the" script of the liturgy, but when I do so, it is the scripts or performance plans of these liturgies that I typically have in mind. Finally, liturgical reenactment comes in different varieties. As will soon be evident, I focus on, not reenactment in all its forms, but those cases in which the performance of speech acts plays a fundamental role. The performance of these acts, I'll claim, plays an important role in the formation and sustenance of traits of character important to the moral and religious life.

Illuminating the way in which participating in the liturgy can contribute to the formation and sustenance of an agent's narrative identity requires some stage-setting. I dedicate the first section of the chapter to this stage-setting and then draw upon this work in the second and third sections to directly address our topic.

I. Liturgical Immersion

The vesperal service of Holy Tuesday begins with a reading from the Gospel of Matthew that presents the story of the woman who the tradition often identifies as a prostitute, Mary of Bethany.[1] According to the scriptural narrative, in the middle of a dinner party, Mary anoints Jesus with "very precious ointment," pouring the oil on Jesus's head and wiping his feet with her tears and hair (Matt. 26:7; Luke 7:37–38; John 12:3). In the services that follow on Holy Wednesday, the hymnody embellishes Mary's story by putting words in her mouth, at one point having her address Jesus with this request: " 'Behold me sunk in sin, filled with despair . . . yet not rejected by your love. Grant me, Lord, remission of my sins and save me . . . O merciful Lord who loves human kind, deliver me from the filth of my works.' "[2] At a different point, the hymnody shifts from the perspective of Mary to that of the assembled, having them sing words attributed to Mary: "Like the Harlot I fall down before you, Christ my God, seeking to receive forgiveness; and instead of ointment I offer you the tears of my heart. Take pity on me, Saviour, as you had on her, and grant me the remission of my sins. For I cry like her to you: Deliver me from the filth of my deeds."[3] What then occurs in the service is that the people file forward to venerate the icon of Christ, kissing it in the way that Mary is said to have kissed Jesus himself on the occasion of the dinner party. Immediately after venerating the icon, each is anointed with oil by the priest.

The sequence of liturgical actions just described is interesting on several levels. Notice, first, that some of the actions performed by the assembled are ones in which they reenact elements of the embellished scriptural narrative by both imitating and repeating actions attributed to Mary, such as her cry to be delivered from her misdeeds. Other liturgical actions, such as when the people venerate the icon of Jesus, imitate Mary's actions but do not repeat them. Instead, they reenact these actions by employing props, in this case an icon of Jesus, to perform actions that are supposed to represent Mary's action of kissing Jesus. It is clear, moreover, that the priest's action of anointing the

[1] The identification is contentious. For a discussion, see Stump (2010), chapter twelve. 12.

[2] *The Lenten Triodion* (Mary and Ware 2002, 539, 537). I have modernized the English used in the translation. In what follows, I will refer to this work as *LT*.

[3] LT, 695–96.

people is supposed to connect with Mary's own anointing of Jesus, although it is not exactly clear how. One interpretation is that the rite is a commemoration of her action, a fulfillment of Jesus's words that what she had done "will be told in memory of her" (Matt. 26). The "telling" in this case would not simply be an oral or a written presentation of her story, but also a dramatic and creative extension of it, in which the priest anoints the people as she anointed Jesus.[4]

Mary is only one of a cast of characters whose actions are reenacted in the context of the liturgy. On the Sunday of the Publican and the Pharisee, for example, the people listen to the Gospel story of the Publican and the Pharisee and sing: "In our prayer let us fall down before God, with tears and fervent cries of sorrow, emulating the Publican in the humility which lifted him on high . . . Let us utter the words of the Publican in the holy temple 'God be merciful.' "[5] Likewise, on the Sunday of the Prodigal Son, the people listen to the story of the Prodigal, and sing, "With the words of the Prodigal I cry aloud: I have sinned, O Father; like him, receive me now in your embrace and reject me not . . . Accept me in repentance and make me as one of your hired servants."[6] In both cases, by performing speech acts of certain sorts, the people reenact parts of the Gospel story by imitating and repeating actions attributed to its characters.

I call the figures of the Publican and the Pharisee "characters" advisedly, as neither is presented in the scriptures as being an actual historical figure. Each is instead a fictional character that Jesus employs in his parables. An interesting feature of these last examples of liturgical reenactment, then, is that when the people imitate and repeat the words of the Publican and the Prodigal, they are engaging with, not a historical narrative, but a fictional one. The case of Mary of Bethany is somewhat different, since she is not presented in the scriptures as a character in a parable. Still, it is worth stressing that when the assembled imitate and repeat Mary's words, they repeat, not Mary's actual words (the scriptures do not present her as saying anything when she anoints Jesus), but those attributed to her by the church's hymnody in its effort to embellish the scriptural narrative. In all the cases we've considered, then, liturgical reenactment involves reenacting either the actions of fictional characters or act types attributed to non-fictional characters that (for all we reasonably believe) they did not perform.

Let me, for ease of reference, call actions of both these sorts *fictive actions*. Given our topic, a natural question to raise is how we should understand the

---

[4] Another interpretation suggested by the liturgical script is that the act of being anointed answers a call issued by Mary of Bethany: "The Harlot washed your pure and precious feet with her tears, and she urges all to approach you and receive the remission of their sins. Unto me also grant her faith, O Saviour, that I may cry to you: Before I perish utterly, save me, O Lord" (*LT*, 375).

[5] *LT*, 107, 108

[6] *LT*, 116, 118.

activity of reenacting fictive actions. Are the reenactments themselves best viewed as a species of make-believe behavior in which the people pretend to be Mary of Bethany, the Publican, or the Prodigal? Or are they better described as cases in which the participants pretend to be present in their own person at, say, the event (whether real or fictional) in which Mary anoints and kisses Jesus or when the Prodigal returns home to his father?

There are powerful trends in both philosophy and theology that would recommend that we view liturgical reenactments of both sorts as dramatic performances and, hence, behavior in which we engage in make-believe.[7] According to these views, when the assembled engage in liturgical reenactment, they should be understood to pretend that they are characters in the scriptural narrative or to be present in their own person at the events depicted by the narrative, much in the way that you or I might pretend to be Mary if we were to stage a dramatic reenactment of the Gospel of Matthew.

I do not wish to broach any empirical claims about what it is that the assembled are doing or take themselves to be doing when engaging in liturgical reenactment. I do want to claim, however, that like any script or score, the liturgical script calls forth responses of certain types from participants in the liturgy. Among the responses *not* ordinarily called forth by the liturgical script, I want also to claim, is that of engaging in make-believe behavior. If I am right about this, the response called forth when the assembled venerate the icon of Christ is not that of pretending to be Mary of Bethany (or to be present in one's own person in Mary's circumstances). Similarly, the response called forth when the assembled imitate and repeat the behavior of the Prodigal is not that of pretending to be the Prodigal (or to be present in one's own person in the Prodigal's circumstances). In the next section, I offer an argument for this assessment. For now I wish to emphasize two points.

The first is that engaging with narratives—whether they be fictional or not—ordinarily requires imaginative activity on the part of those who attempt to understand or interpret them. Among other things, when we try to understand a narrative, we must fill in its gaps, project possible explanations for what occurs, or creatively interpret its happenings in the light of relevant background knowledge, and so on. But we should not, I believe, equate imaginative activity with pretense; to imaginatively extrapolate from a narrative needn't be to engage in any sort of make-believe behavior. The second point is that from the fact that important parts of a narrative are presented as and recognized to be fictional, we cannot infer that imaginatively engaging with that narrative is itself a matter of engaging in make-believe. When one finds oneself imaginatively engaged by a story—even to the point of being swept away by it—one needn't be engaging in any sort of pretense or game of make-believe. There is

[7] See Walton (1990) and the references in Wolterstorff (1990), for example.

such a thing as imaginative engagement with fictional narratives that is not pretense.

Let us suppose for present purposes that the assessment just offered is correct: when the script directs the assembled to reenact the behavior of the Publican or the Prodigal by imitating or repeating it, it ordinarily calls forth not the activity of pretending to be the Publican or the Prodigal (or pretending to be present in one's own person in their circumstances) but of imaginatively engaging with the stories that involve these characters. How, though, should we understand the sort of imaginative engagement that the liturgical script calls forth if it is not a species of make-believe?

The issue calls for a discussion unto itself but let me briefly gesture at what I believe is the correct answer.[8] There are, in principle, untold ways in which one could approach a work, such as a novel, that presents a narrative; one could approach it for the purpose of uncovering the author's religious views, political allegiances, sexual preferences, or the like. Here, however, is a way to approach such a work with which we are familiar and on which many other approaches are parasitic: one imaginatively immerses oneself in the narrative presented by the work.

Think of the activity of immersing oneself in a narrative work (that is, a work that presents a narrative) as having two primary components: first, it involves attending to the content of the narrative of the work (what it is communicating) and the properties of that content (such as how its various elements hang together) in such a way that, in engaging with the work, one prioritizes attending to this content. This means that, while immersing oneself in a narrative work is compatible with attending to features of the presentation of its narrative, such as the author's word choice or use of certain grammatical constructions, doing so is for the purpose of better attending to the content of the narrative itself. So, when reading, for example, one can momentarily marvel at the use of a striking metaphor, asking oneself why the author would use it in this context. But the point in doing so is better to engage with the content of the work in which the metaphor is being used.

Immersing oneself in a narrative work involves more, however, than simply attending to its content in such a way that one screens off or fails to prioritize certain features of its presentation, such as the author's word choice. It is also to take up a certain kind of vantage point with regard to the narrative presented by the work. Explicating the phenomenon of taking up a vantage point is not easy, but the basic idea is that when immersing oneself in a narrative work, one imaginatively enters its narrative by situating oneself within it.[9] In taking up such a vantage point, then, one does not take up the stance

[8] I have developed the view in considerably more detail in Cuneo (2014a).

[9] Harris (2000) offers some interesting empirical data that supports this way of thinking about immersion.

of a dispassionate observer or critic. Rather, one attends to its content in such a way that its characters and events loom large in one's consciousness, and one becomes emotionally engaged to some significant degree with these characters and events. Still, in so immersing oneself in the narrative of the work, one does not take oneself to be a character in the work (or be present in one's own person at the events described in the work). Characters in a work do not, after all, attend to the ways in which the work of which they are a part is presented.[10]

The positive proposal I wish to make is that the liturgical script ordinarily calls forth behavior similar to that in which we engage when we immerse ourselves in a literary work. When the script of the service of Holy Wednesday, for example, calls for the assembled to imitate and repeat the speech acts attributed to and the behavior of Mary of Bethany, they are to immerse themselves in Mary's story. Or, to state the phenomenon from the opposite angle, when the script directs the assembled to imitate and repeat the speech acts attributed to and behavior of Mary, it calls forth the activity of allowing the assembled to be absorbed by Mary's story, engaging with its content and allowing her character and actions to loom large in their consciousness and emotional life when they reenact her behavior.

II. Identification

I started our discussion by raising the question of why the scripts of the ancient liturgies go beyond directing the assembled simply to read and listen to segments of the core Christian narrative to also reenact some of the events it depicts. In my initial attempt to address this question, I assumed that it would be helpful to get a clearer sense of what liturgical reenactment is, the character of the sort of activity that is called forth when the liturgical script directs the assembled to do such things as imitate and repeat the actions of Mary of Bethany or Jesus. The question on which we need to make progress is the contribution that liturgical reenactment so understood is supposed to make to the moral and religious life. Let me take a few more steps in toward that goal by saying something more about the phenomenon of reading narrative works, canvassing what some philosophers have said about the ways in which reading literature is supposed to contribute to the moral life and the formation of character.

In the introduction to her collection of essays *Love's Knowledge*, Martha Nussbaum points out that moral philosophers have virtually ignored the role

[10] Cuneo (2014a) expands upon these points, noting that immersion typically requires moving back and forth between attending to the content of a work and ways in which that work is presented.

of literature in the moral life.[11] Nussbaum contends that this oversight is not trivial, for literature—and Nussbaum has in mind the novel in particular—is especially well-suited to contribute to the expansion and refinement of moral understanding. What literature does, Nussbaum suggests, is present us with rich descriptions of characters, their traits, and their often complicated predicaments, allowing us to emotionally engage them and thereby expand our powers of moral judgment, refine our abilities of ethical discernment, deepen our ethical understanding, and have access to alien points of view. In this regard, Nussbaum contends, these works are importantly different from those of philosophy. In a treatment of the virtues, a philosopher such as Aristotle or Aquinas might offer an accurate analysis of a particular virtue. But given its abstract nature, such an analysis could (for most of us) gain only limited traction in the moral life; we would gain very little appreciation for its lived character. By contrast, "showing" a virtue by means of a narrative that displays its dynamics over time—so Nussbaum maintains—offers us a much better sense of what the virtue is and how we might recognize it. In this way, the sorts of narratives regularly presented in literature make a unique and important contribution to the moral life; they often accomplish what works from other genres—and, indeed, ordinary life—cannot.

When all goes well, immersing oneself in liturgical action might expand and refine our powers of moral understanding in much the same way that engaging with literature does, at least if philosophers such as Nussbaum are right. But it would not do so because those segments of the liturgical script concerned with liturgical reenactment present us with rich and detailed descriptions of characters and their predicaments. The reason is that although these segments of the liturgical script engage with the scriptural narrative, they do so largely by re-presenting and embellishing portions of this narrative. More exactly, these segments of the script are largely a selective and stylized appropriation of this narrative that presupposes an intimate familiarity with it. When, for example, the script for the services of Holy Friday imaginatively weaves together the stories of Joseph of Arimathea and Mary Theotokos, presenting these stories from the first-person perspective of the characters, it does not follow them through an extended period of time, offering rich and detailed descriptions of these characters and the predicaments they face. Rather, it offers snapshots of their lives, creatively blending and embellishing their stories by, among other things, highlighting some of their important characteristics, such as their courage in conditions that appear to them hopeless.[12] But if the contribution that liturgical reenactment makes to the moral and religious life is not that

[11] Nussbaum (1992); Carroll (2001, part 4; 2010, parts 4 and 5) also have some of the best discussions of the topic of which I'm aware.

[12] Eric Auerbach's (2003) classic work *Mimesis* contrasts Greek narrative and biblical narrative on precisely this point: biblical narratives don't fill in the background in the way that Greek narratives do.

of imparting moral understanding by the presentation of rich and nuanced narratives, how should we understand it?

The answer, I believe, lies in what I will call the self-reflexive character of the liturgical script. As the passages I have quoted from the liturgical script will have already indicated, the liturgy has the striking feature of casting much of its hymnody in the first-person. This technique is employed in three different though compatible ways.

In some places, the technique is employed to recast stories, which in their scriptural presentation are themselves not narrated from the first-person perspective, into hymns that are narrated from the first-person perspective of their characters, such as when the hymnody presents elements of the Genesis story from the perspective of Adam:

> The Lord my Creator took me as dust from the earth and formed me into a living creature, breathing into me the breath of life and giving me a soul. . . . He honoured me . . . making me companion of the angels . . . In my wretchedness I have cast off the robe woven by God . . . and I am clothed now in fig leaves and in garments of skin.[13]

Elsewhere, the script presents hymnody from the perspective of Jesus, the so-called second Adam:

> I who am rich in Godhead have come to minister to Adam who is grown poor. I who fashioned him have of mine own will put on his form. I . . . have come to lay down my life as a ransom for him.[14]

Call the technique employed in this and other passages *narrative recasting*. If the point of these particular examples of narrative recasting were simply a matter of getting the assembled to pretend to be Adam or Jesus, their function would be more or less transparent. What better way to pretend that one is Adam or Jesus than to take up his perspective? I have claimed, however, that we ought not to understand the liturgical script to call for the assembled to pretend that they are characters, such as Adam and Jesus. If I am right about this, when the script directs the assembled to sing hymns from Mary or Jesus's perspective, something else is going on. That something else, I would say, consists in the assembled *playing the roles* of Adam and Jesus—where playing the roles of these characters consists not in pretending to be them but (in part) taking up their perspectives by speaking in their voices.[15]

In other places, the liturgical script directs the assembled not to recite stories narrated from the first-person perspective of the characters, but to quote

---

[13] *LT*, 168.

[14] *LT*, 513.

[15] It is not easy to specify exactly what it is for an agent to play a role. I say more about the matter in Cuneo (2014a).

and appropriate passages from these stories, employing them to perform speech acts that express the first-person perspective of those assembled:

> [L]et us emulate the groaning of the Publican and, speaking to God with warm tears, let us cry out: "O you who loves humankind, we have sinned. In your compassion and pity, be merciful and save."[16]
>
> Like the Thief I cry to you "Remember me"; like Peter I weep bitterly; like the Publican I call out "Forgive me, Saviour"; like the Harlot I shed tears. Like the woman of Canaan I cry to you, "Have mercy on me, Son of David." Like the woman with an issue of blood, I touch the hem of your garment. I weep as Martha and Mary wept for Lazarus.[17]

Call this activity that the liturgical script calls forth *indexical appropriation*. In the cases before us, indexical appropriation consists in embedding various speech acts, such as requests that incorporate first-person pronouns ("Remember me" and "Forgive me, Saviour"), in other speech acts, such as assertions ("Like the Thief I cry to you") that also incorporate first-person pronouns. Unlike the examples of narrative recasting offered above, indexical appropriation consists not in playing the role of a character but allowing biblical persona to function as models—these models often functioning as ideals whose characteristics we try to approximate in our own behavior.

Finally, in other places, the liturgical script employs the first-person pronouns in what is perhaps an even more unusual way. Under this use, the pronoun functions not so much as a marker of who is doing the addressing as an indicator of who is the *object* of address. When the pronoun is used in this way, the liturgical script is self-referential in the sense that, when an individual conforms to it in the context of liturgical performance, she is the object of her own address. What is more, these uses of the first-person pronouns often figure in a certain type of construction in which the speaker refers to him- or herself comparatively, as being like or identical with one or another character, such as when the assembled compare themselves to the Prodigal Son:

> I am become the Prodigal Son, and having wasted my riches I perish now from hunger. Beneath your protection I seek refuge, O loving Father: accept me as you have accepted him. Make me a sharer at your table, that I may cry to you: Before I perish utterly, save me, O Lord.
>
> I am the prodigal: conceived in sin, I dare not look up to the height of heaven. But trusting in your love for human kind, I cry: God be merciful to me and save me.[18]

---

[16] *LT*, 108.
[17] *LT*, 245, 263.
[18] *LT*, 375, 668.

Call this use of the singular first-person pronoun *comparative self-address*. Unlike narrative recasting, comparative self-address is not (in the paradigm case) a matter of playing a role. And unlike indexical appropriation, it is not (in the paradigm case) allowing a character to function as a model. Rather, I would say it consists in treating a figure function as a *type*, of which the assembled declare that they are examples. In the passages just quoted, for example, the Prodigal Son functions in just this way.

The overarching question we are pursuing concerns the contribution that liturgical reenactment is supposed to make to the moral and religious life. An important clue to understanding that contribution, I have suggested, is the liturgical script's use of the first-person pronouns. Let me now move beyond noting the various uses to which the script puts these pronouns and hazard a general suggestion as to what their functions might be in the context of liturgical performance. Having offered this suggestion, I will qualify it in some important respects.

Imagine that you and I are very different people: you are a thief, and I am not. Or somewhat differently, suppose that you are a collaborator with an oppressive foreign power that is occupying our country, and I am not. Between you and me there is likely to be considerable psychological distance. That is, given our different stations, it is likely that you and I have rather different histories, temperaments, values, commitments, and the like. Suppose, though, it is important for me to understand you better, say, because we find ourselves invested in a joint project whose success matters to both of us. In order to understand you better, I could engage in at least two activities.

I might attempt, in the first place, to close the psychological distance between us by taking up your point of view on some matter on which I find your views puzzling or alien. Or, somewhat differently, I might attempt to identify with you, locating points of similarity between you and me, where these points of similarity could be the basis for mutual enjoyed recognition, lasting social bonds, or unified action, such as when we realize that we share a deep interest in jazz and make plans to perform some of our favorite pieces with one another.[19] A moment's reflection reveals that these activities are not identical. I could close the psychological distance between you and me to some significant degree by appreciating how you see things without identifying with you in any appreciable sense, since your ways of seeing the world might still seem too alien. Similarly, I could identify with you without having to close any appreciable psychological distance between us, since there might be none to speak of.

I take it that closing the psychological distance between oneself and another is often for the purpose of identifying with that other person. At any rate, in

[19] A third activity would be to imagine what it would be like for me to be in your position.

what follows I shall understand it to have that function and will speak simply of the activity of identifying with another—where identifying with another is understood typically to incorporate the activities of both closing the psychological distance between and identifying (in the sense specified) with that person.

The proposal I want to develop is that the liturgical script's use of narrative recasting, indexical appropriation, and comparative self-address is to facilitate the identification with the array of characters presented in the liturgical script. To appreciate the point, begin by observing that the liturgical text presents us with an array of characters that many of us are likely to find deeply alien in important respects: criminals (the Thief), collaborators with oppressive foreign powers (the Publican), betrayers (Peter), wastrels (the Prodigal), prostitutes (Mary of Bethany), the hysterical (the woman of Canaan), and the impure/sick (the woman with an issue of blood). Notably, however, the script directs the assembled not to distance themselves from these characters because they are vivid examples of the unsavory and the unclean but to imitate and repeat their actions and words to the point of identifying with them, sometimes in the strongest of terms, such as when it directs the assembled to assert such things as "I am the Prodigal."

That identification is in fact the activity called forth by the liturgical script is, I think, especially evident in the cases of comparative self-address that we considered. But it is also evident in the cases of indexical appropriation that we canvassed, such as this passage:

> Like the Thief I cry to you "Remember me"; like Peter I weep bitterly; like the Publican I call out "Forgive me, Saviour"; like the Harlot I shed tears. Like the woman of Canaan I cry to you, "Have mercy on me, Son of David." Like the woman with an issue of blood, I touch the hem of your garment. I weep as Martha and Mary wept for Lazarus.[20]

This passage helps us to see the self-constituting character of liturgical reenactment: in performing the speech acts of requesting forgiveness and mercy by using the words of the Publican and the woman of Canaan, the assembled do not simply express a desire to emulate these characters; they also *bring it about* that in performing these speech acts a range of similarities now hold between them and these characters—these similarities being of such a kind that they can contribute to the project of identifying with these characters. To the question, in what way are you like the Publican? it could fairly be responded by the assembled: I am like the Publican in imitating and following his example to the point of even using his words. Imitation, in this case, begets identification.

---

[20] *LT*, 245, 263.

The liturgical activity of identifying with the range of characters presented in the liturgy is intriguing if only because traditional religious traditions are often intent on carving out and ratifying strict social roles for their adherents. In many such traditions, men are allowed to perform certain actions, but women are not. The ritually clean are permitted to participate in certain rites, but the ritually unclean are not. In the passages that we have considered, exactly the opposite dynamic appears to be at work; the script invites men to identify with characters who are women; women, with characters who are men; the innocent, with characters who are criminals; the faithful, with characters who are betrayers; the responsible, with characters who are wastrels; the healthy, with characters who are chronically sick; the stoical, with characters who weep; the emotionally stable, with characters who are emotionally unstable, and so forth.

The intended effect seems (at least in part) to be twofold. On one level, it is to destabilize the self to some degree, inviting participants in the liturgy to expand their own self-conceptions in ways that, for many, are hardly going to seem like second nature. One could, in fact, look at the activity of identifying with the cast of characters presented in the liturgy as an exercise of autonomy of a certain kind in which agents revise certain self-understandings of themselves or call into question roles into which they have been scripted. On another level, however, the script seems intent on unifying the self-conceptions of the participants around certain ethical and religious ideals that these characters share, such as their willingness to repent of their failures and their readiness to acknowledge that they do not have it all together. In this regard, participating in liturgical reenactment of this sort has a pronounced leveling effect: whoever you might be, and whatever your station—so the script seems to say—there is considerable common ground between you, me, and these characters whose actions and behaviors we imitate and repeat. Interestingly, unlike the basis of the sort of friendship that Aristotle lauded, the common ground does not so much consist in acknowledging shared virtues or accomplishments as deep and regrettable flaws, ways in which we are fragile and fail.

Let me now add a qualification to these points about identification, which I mentioned in passing a few paragraphs back. To this juncture, I have been speaking in general terms about the function of the liturgical script's use of the first-person pronouns, suggesting that they are employed to invite the assembled to identify with these various biblical persona. But speaking in such general terms risks distortion, since terms such as "I" and "we" are indexicals and, as such, do not have a unique content that determines their reference. All else being equal, when used by me, the term "I" refers to me; when used by you, the term refers to you. Indexicals, then, have what are commonly called *characters*, which are, roughly, rules or functions, from contexts—such as my uttering the term "I" on some occasion—to contents—in this, case me. The character of the

first-person indexical, then, will determine on any given occasion of use who it picks out or refers to.

I call attention to this point about the nature of indexicals because it is important to see that in its ample use of first-person pronouns, the liturgical script allows for a great deal of flexibility concerning what is said or accomplished by the participants' use of these pronouns. Depending on who you are and what stage of your life that you are in, your uttering sentences in a liturgical context that repeat those attributed to the Thief or the Prodigal Son might mean something very different or play very different cognitive roles from my uttering those same sentences in the same context. Given your history, uttering the sentence "I am the Prodigal" might function as an invitation to search for similarities that might not be obvious or to close the psychological distance between you and the Prodigal so that you can better understand and reconcile with those who are like him in obvious respects. In contrast, given my history, uttering that same sentence might function to acknowledge a past that I would rather forget but that is important for me to come to terms with.

I would add to this that the activity of identifying with characters in the context of liturgical performance can be fraught with moral risk, as it can be an occasion to distort one's self-conception or trigger tendencies that are on the whole harmful to oneself, such as engaging in self-abasing behavior when what is needed most is assurance of one's worth. We have, then, another reason to think of the liturgical script as calling forth certain types of responses, not from all participants, but from only those who stand to make progress in the moral and religious life by immersing themselves in the activity of liturgical reenactment. It goes without saying that whether one is such a person might require reflection and good judgment.

To this point, I have been developing the thought that, like engaging with literature, liturgical reenactment can make important contributions to the moral and religious life, helping us to realize certain ethical and religious ideals in the activity of identifying with the array of characters presented in the liturgical script. Recall, though, that my claim at the outset of the chapter was stronger than this, for I said that a dominant aim of liturgical reenactment is to contribute to the construction of a narrative identity. And, it could rightly be observed that identifying with someone in some respect is not perforce to construct a narrative identity. I could, for example, discover that you and I have a shared love of jazz, and we might identify with one another because of this. But when it comes to telling a story of my life, this similarity may not be important enough to figure in my story. It is not part of my narrative identity.

That noted, it strikes me that attention to the cases of comparative self-address that we've considered, such as when the assembled assert "I am the Prodigal," strongly suggests that the liturgical script is calling forth more than simply identifying with characters such as the Prodigal. It is telling, after all, that in the passages that I have quoted, the identification is complete; the

claim is not that I am like the Prodigal in some respect or other but that I *am* the Prodigal. I am the one who has thrown away his birthright and squandered the gifts given to him. As such, the identification appears to be of such a kind and importance that it is an answer to the question, "who are you?" which is exactly the sort of question that the telling of one's story is supposed to address. In fact, I think we should be open to the possibility that what the liturgical script is calling forth when it directs an agent to reenact the Prodigal's actions and engage in comparative self-address is for that agent to narrate or construct his or her own story in the context of the liturgy. If that is right, the activity of liturgical reenactment can be itself a kind of storytelling, a telling of one's own narrative. That is, such an activity can consist in articulating who one is or aspires to become, which would be another striking example of the self-constituting character of the liturgical action, since engaging in such an activity one would, in a certain range of cases, bring it about that one's narrative identity has a particular contour, which it shares with the Prodigal's. I would add that the telling of one's story in the context of the liturgy needn't be de novo. For many, it will be a retelling of one's story that perhaps amends, augments, or corrects a narrative that one is prone to tell in other contexts. At any rate, if these suggestions are along the right lines, we can identify yet another way in which narrative functions in the context of the liturgy: when people assemble for the liturgy, they not only read and listen to narratives but are also called to construct them while engaging in the liturgy.

III. Commitment

The overarching question that I have been pursuing is the contribution that liturgical reenactment is supposed to make to the moral and religious life. The clue to answering this question, I've suggested, lies in the self-reflexive character of the liturgical script, as it is this feature of the script that helps us to locate the sorts of activities the script calls forth. Prominent among the activities called forth by the script, I have claimed, is that of identifying with characters in the core Christian narrative, where this identification should be understood as contributing to the construction of a narrative identity.

I have, however, stopped short of broaching any empirical claims to the effect that liturgical reenactment is likely to contribute to the construction of a narrative identity. For my proposal is unapologetically normative: the construction of a narrative identity is among the activities that the script calls forth. I want now to suggest that the answer to our overarching question is normative in another distinct sense.

The cases of reenactment in which I've been interested involve the performance of speech acts of various sorts in which one appropriates the words of characters presented to us in the liturgical script. Fundamental to the performance of

any speech act, however, is that the speaker commits himself to the world being a certain way. In asserting, I commit myself to the world being thus and so (and to my believing it to be thus and so). In promising, I commit myself to acting in a certain way (and believing that I can act that way). In commanding, I commit myself to having the authority to direct your actions in certain ways (and believing myself to have such authority). In my view, it is the normative alteration that occurs upon committing oneself in these ways that accounts for (at least in part) why it is that certain sentence utterances count as speech acts such as asserting, promising, and commanding. Speech, under this approach, necessarily involves normative transformation of such a kind that a speaker acquires rights, responsibilities, and obligations vis-à-vis her audience, and vice versa.[21]

Taking account of the ways in which speech is normative allows us to more adequately address questions that I raised earlier in this discussion. One question—the question that I raised at the very outset—is why the scripts of the ancient liturgies go beyond directing the assembled to read and listen to the presentation of events that compose the core Christian narrative to reenacting them. The answer is that in reading or listening to a narrative, one does not thereby commit oneself to anything. But when the assembled engage in reenactment of the sort called forth by the liturgical script, they perform speech acts of various sorts that commit them to being certain ways, such as being like (or aspiring to be like) the Thief, the Publican, or the Prodigal.

A second, closely related question is how liturgical reenactment might make a distinctive contribution to the moral life, different from reading narrative works such as novels. The answer is that while reading a novel can be morally transformative, it is not an activity that as such calls forth the activity of committing oneself to anything. Liturgical reenactment of the sort called forth by the liturgical script is different since it belongs to the essence of this activity that the assembled commit themselves to ethical and religious ideals of various sorts, including being like characters such as Mary of Bethany. Of course we have excellent empirical evidence that in a large range of cases these ethical ideals do not "take." But when they do not, the failure is of a distinct kind; it is a failure to live up to standards that have not been imposed on one by another but to which an agent has committed herself in the context of liturgical action.[22] If this is so, and liturgical reenactment is as I have described it, then such reenactment is doubly normative: it is an activity not only called forth

[21] Cuneo (2014b) defends this position. This view, which I call the *normative theory of speech*, takes its inspiration from Alston (2000); Brandom (1998); Searle (1969); and Wolterstorff (1995).

[22] In the Divine Liturgy, immediately before partaking of eucharist, the assembled pray: "May the communion of your holy mysteries be neither to my judgment nor my condemnation, O Lord, but to the healing of soul and body." One way to understand this prayer is that it expresses the desire that the partaking of eucharist not be out of step with ideals to which one has otherwise committed oneself in the context of the liturgy in such a way that partaking of the eucharist proves to be an exercise of self-condemnation, a failure to live up to these ideals to which one has committed oneself.

by the liturgical script but also one in which the assembled commit themselves to various moral and religious ideals, including being like (or aspiring to be like) the characters presented in the liturgical script.

Having noted the ways in which liturgical reenactment is normatively transformative, we are now better situated to address yet another issue that I left hanging. The issue is why we should understand liturgical reenactment to be a species of not make-believe behavior but immersion in which one imaginatively enters into the narrative presented by the liturgical script. We are better situated to address this question because it should be evident that when one engages in pretense of a sort in which one pretends to be a character, one does not thereby commit oneself to the world or oneself being a certain way. To illustrate, suppose I were to pretend to be Mary of Bethany, repeating the words attributed to her in the liturgical script: "Behold me sunk in sin, filled with despair . . . yet not rejected by your love. Grant me, Lord, remission of my sins and save me . . . O merciful Lord who loves human kind, deliver me from the filth of my works." Were I to repeat Mary's words in the context of the liturgy while pretending to be her, I would not thereby have committed myself to being sunk in sin, filled with despair, or anything of the sort. Neither would I have requested anything of Jesus. I would have merely acted the part of Mary of Bethany much in the way that I would if I were acting a part in a theatrical presentation of the book of Matthew. If I am right to suggest that the heart of liturgical reenactment consists in *committing* oneself to being certain ways, however, then I believe that we have decisive reason to reject an account of liturgical reenactment according to which it consists in make-believe behavior. For if it did, then liturgical reenactment could not be normatively transformative—at least in the ways with which we have been concerned. Its aim would not be to transform the self by way of committing oneself to certain ethical and religious ideals.

There is one remaining observation that deserves to be voiced, and that is the communal nature of liturgical reenactment. For, although I have not emphasized the point, liturgical reenactment is an activity that is performed not in isolation but in the context of a community dedicated to the realization of certain moral and religious ideals. When all goes well, one is aware that the liturgical script calls one's fellow participants to commit themselves to the very ideals to which it calls you to commit yourself. Moreover, your fellow participants are aware that the script calls you to commit yourself to the ideals to which it calls them to commit themselves. Mutual recognition, then, is yet another dimension of liturgical reenactment and it deserves to be the subject of further reflection. It is a topic of sufficient richness, however, that exploring its role will have to await another occasion.[23]

[23] Thanks to Brad Cokelet, Christian Miller, Nick Wolterstorff, members of the University of Vermont Ethics Reading Group, and an audience at the Character Project conference at Wake Forest University. Work on this paper was made possible through the support of a grant from The Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this

References

Alston, William. 2000. *Illocutionary Acts and Sentence Meaning*. Ithaca, NY: Cornell University Press.

Auerbach, Erich. 2003. *Mimesis*. Princeton, NJ: Princeton University Press.

Brandom, Robert. 1998. *Making It Explicit*. Cambridge, MA: Harvard University Press.

Carroll, Noël. 2001. *Beyond Aesthetics*. Cambridge: Cambridge University Press.

Cuneo, Terence. 2014a. "Liturgical Immersion." *Journal of Analytic Theology* 2: 117–39.

Cuneo, Terence. 2014b. *Speech and Morality*. Oxford: Oxford University Press.

Harris, Paul. 2000. *The Work of the Imagination*. Malden, MA: Blackwell.

Mary, Mother, and Kallistos Ware. 2002. *The Lenten Triodion*. South Canaan, PA: St. Tikhon's Seminary Press.

Nussbaum, Martha. 1992. *Love's Knowledge*. Oxford: Oxford University Press.

Searle, John. 1969. *Speech Acts*. Cambridge: Cambridge University Press.

Stump, Eleonore. 2010. *Wandering in Darkness*. Oxford: Oxford University Press.

Walton, Kendall. 1990. *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Wolterstorff, Nicholas. 1990. "Remembrance of Things (Not) Past." In *Christian Philosophy*, 118-161, edited by Thomas Flint. Notre Dame, IN: University of Notre Dame Press.

Wolterstorff, Nicholas. 1995. *Divine Discourse*. Cambridge: Cambridge University Press.

publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation.

27 }

Cultivating Virtues through Sartorial Practices: The Case of the Islamic Veil in Indonesia

Elizabeth M. Bucar

This chapter provides an overview of the role of bodily practice in moral formation by exploring the manner in which public Islamic veiling[1] cultivates a woman's character in the Indonesian context. My contribution to this volume draws from a month of fieldwork in Java, Indonesia, in 2011, during which I conducted interviews, organized focus groups, took pictures, and read advice literature as part of a research project designed to be substantially different from the usual scholarly work on the Islamic veil. The focus of this research was on the interactions between ethics and aesthetics in the practices of *fashion veiling*—wearing the Islamic dress while following fashion trends—and on the manner in which aesthetic value might be perceived as moral or immoral. This pious fashion raises various ethical questions, for example: Can a woman embody the virtue of modesty when her intent is to wear a style that flatters her? Can certain forms of Islamic dress actually increase a woman's physical attractiveness to others? And, what happens socially to the power of veiling when it suddenly becomes "in vogue"? My aim in a larger research project is to extend an interdisciplinary, comparative conversation of how visual culture orders, expresses, and produces virtue.

The concept of virtue has been at least tangentially deployed in some existing scholarly treatments of the veil. For instance, scholars have used virtue to explain why veiling has continued to be a popular practice in Islamic communities

[1] I use the term "Islamic veil" instead of "Muslim veil" because Muslims do not construct the veil's meanings in isolation. Non-Muslim ideas about Islam, even erroneous ones, have also impacted the discourse, practice, and meaning of the Islamic veil. The veil is therefore not an object that purely reflects Muslim culture, politics, or religious ideals. What holds these sartorial practices together at least in the contemporary world is that they are tied to Islam both by the practitioners and by those who observe them.

despite the ways in which it clashes with other local social norms; many Muslims claim the practice is necessary to create virtuous dispositions, as well as virtuous actions (e.g., Mahmood 2005). A virtue approach has also been used indirectly to explain why veiling remains a contentious issue today by highlighting competing beliefs about what virtues are important (such as modesty and tolerance), and what public practices are seen as too "ostentatious" to respect beliefs about civic virtues (e.g., Scott 2007).

I have two aims in the chapter. First, I use a virtue framework more directly to analyze Muslim sartorial practices. Second, I consider the case of Islamic dress in order to rethink virtue ethics more generally. Specifically, I suggest that understanding modest dress is a way to begin to construct a form of feminist virtue ethics that accounts for the role of gendered practices in the cultivation of character.

This chapter is organized into three sections. The first is an overview of Qur'anic directives to veil. This provides textual context for why many Muslims consider a women's dress an important means to and expression of her good character. In the second, results from my study of veiling in central Java are summarized, with a focus on the role of fashion during a women's transition to wearing the veil, advice literature about how to veil, and debates over flattering veiling and spiritual beauty. The third section considers themes related to virtue that arise in Indonesia fashion veiling and uses these to propose ways in which women's modest dress might encourage us to rethink assumptions about the cultivation of character. A brief conclusion suggests directions for future research.

Scriptural Context for Veiling

When asked why they cover their head, body, or face, most Muslim women reply that they are following a written directive in Islamic scripture and that they do so in order to make themselves into good Muslim women. Three verses are commonly cited as Qur'anic evidence of the Islamic veil—33:53, 33:59–60, and 24:30–31. However, there is enormous diversity among these verses as well. The terms that get translated as "veil" are different: *hijab* (curtain), *jilbab* (cloak), *khimar* (kerchief). Who is asked to veil, for what reasons, and in front of whom differ in these verses, which complicates what norms are at stake in the practice. Various virtues are related to women's dress, including modesty, purity, honor, righteousness, and faith. But Muslims have often interpreted these verses as requiring specific bodily actions of Muslim women in order to for them to become and be pious.[2]

[2] There are alternative interpretations of the ethical meaning of these verses. For an overview of the main interpretative issues, see Bucar (2012, 28–48).

The verse that connects veiling most explicitly to the cultivation of modesty is 24: 30–31, often referred to as the verse of the *hijab*, which reads as follows:

> Say to the believing men that they should lower their gaze and guard their modesty [*furujahun*]: that will make for greater purity for them: and God is well acquainted with all that they do. And say to the believing women that they should lower their gaze and guard their modesty; that they should not display their beauty and ornaments [*zina*] except what (must ordinarily) appear thereof; that they should draw their veils [*khumur*, sing. *khimar*] over their bosoms [*juyub*] and not display their beauty [*zina*] except to their husbands, their fathers, their husband's fathers, their sons, their husband's sons, their brothers or their brother's sons, or their sister's sons, or their women, or the slaves whom their right hands possess, or male servants free of physical needs, or small children who have no sense of the shame of sex [*'awra*]; and that they should not strike their feet in order to draw attention to their hidden ornaments. And O you Believers! You turn all together toward God, that you may attain Bliss (Ali 2001, 261–262).

In verse 24:30, men and women are asked to *do* certain things in order to *be* a certain way, specifically to "lower their gaze," an action "that will make for greater purity for them." Qur'anic commentators have tended to interpret averting the eyes metaphorically as avoiding desiring what is forbidden. But Muslims also interpret this quite literally, so that modesty understood at its most fundamental level is about how one physically interacts with others. Physical submission[3] (lowering the gaze) is recommended in interpersonal interactions in order to prevent intimacy and to cultivate the virtue of modesty.[4] By not looking at someone directly, we not only do not "flirt," we also maintain bodily distance despite physical proximity. The more a person does this practice, the more it becomes automatic, and the more successful she will be in cultivating the virtue of modesty.

Lowering one's gaze is a bodily action suggested for both men and women, but verse 24:31 requires a second category of bodily actions for women: they should not display their "ornaments." This is most often interpreted as requiring a specific sartorial practice. Since the meanings of both *zina* and *'awra* are up for debate, it is not particularly helpful that in verses 24:30–31, the things a woman must cover with clothing are defined by these terms. Luckily, we are given a few other hints about what proper dress entails in the verse itself. For

[3] It is important to note that in Islam, submission to God is not seen as in conflict with human freedom; rather, in a way that will be familiar to those who work on Christian conceptions of virtue, our submission to God is the very condition of our true freedom.

[4] Elsewhere in the Qur'an (e.g., 33:33), modesty is juxtaposed against *tabarruj*, an overt form of flirtation that involves prancing, flaunting, and embellishing with the goal of attracting attention.

one, we are told Muslim women should cover their *juyub* with *khimar*. The root of *juyub* means an "opening" or "space between." Although it is often translated as bosom, it is technically a woman's cleavage. *Khimar*, a term some commentators gloss as "veil," refers to a specific type of clothing, mainly a kerchief. In other words, the verse tells us that cleavage is part of hidden *'awra* and thus should be covered by something a woman wears. The implication is that in addition to a specific form of physical interaction that is required of men and women (lowering the gaze), women should dress a certain way as well in order to become modest.

The Case of Contemporary Veiling in Indonesia

Despite Indonesia's being the most populous Muslim nation in the world, Indonesian women did not historically cover their heads. In fact, *jjlbab*, as Islamic women's modest dress is called locally, has only become common in Indonesia in the last ten years. On the one hand, the cachet of fashion veiling in Indonesia is tied to a desire to be viewed as a modern cosmopolitan nation: local fashion is read as a sign of successful development (Jones 2007, 216). On the other hand, fashion veiling has also increased the visibility of public piety by literally making veiling more attractive to young Muslim girls. For these reasons as well as others, fashionable *jilbab* style is, for the most part, not only tolerated in Indonesia, it is also respected and admired as the sign of a successful cultivation of character.

Dress codes do vary by region, but most Indonesian women are free to choose whether and how they will veil. This means for women who cover that there is a point at which they transition from not wearing the veil to wearing one, often in high school or college. One way to understand this transition is in terms of identity. When a Muslim woman decides to don the veil, we might interpret this as the moment when her identity shifts from one that is ascribed (I was born a Muslim, raised in a Muslim home) to one that is declared and thus salient (I chose my Muslim identity after consideration of alternative options). In addition, she becomes visibly Muslim in a more tangible way.[5]

During this transition to being covered, women who follow fashion trends before veiling try to maintain their personal style when deciding how to veil. For others, a new interest in fashion coincides with covering. During an

[5] Emma Tarlo has written eloquently on the veil and visible Muslim identity in the United Kingdom (Tarlo 2010). This legibility, and its ethical impact in different political contexts, has enormous potential for work in comparative ethics. For instance, the American woman who puts on the veil for the first time after 9/11 or the French schoolgirl who veils at the risk of expulsion are using visual gestures to mean and do different things within specific contexts.

interview over coffee at Am Plaz, a glitzy mall in Yogyakarta, fashion blogger Rania confided to me,

> I started wearing *jilbab* three years ago, when I was 31, already married, and had three children. At the time, I was thinking a lot about death and I thought I wanted to be a better person. *Jilbab* can't make you a better person, but I knew Muslim women should cover and so I wanted to cover to become a better person. Before I covered I thought that people who were into fashion were materialistic. . . . But when I started to wear *jilbab* I realized that I didn't have anything to wear except t-shirts and jeans . . . For me, *jilbab* was a fashion opportunity.

Rania later describes how her process of learning "fashion skills" began by searching the Internet for styles that she found appealing. She collected images, which eventually became the foundation of her blog: "The things I post on my blog are the things I found interesting or helpful and I want to share them with other *hijabis* [women who wear Islamic dress including a headscarf]."

In addition to blogs like Rania's, a new genre—headscarf advice literature—is consulted by women in the process of selecting a style. This literature gives step-by-step instructions for wrapping, twisting, and pinning headscarves to create different forms; advice about "starter kits" that women should buy, such as the number and color of pins, brooches, and scarves; recommendations for how to match the pattern and color of a headscarf to the rest of the outfit; and guidelines for how to choose a style, emphasizing the importance of wearing the veil in a flattering way.

For instance, women with round or square faces are given the following warning:

> Avoid wearing headscarves or base layers that cover the forehead because they will give the impression that the face is shorter. Avoid wearing a headscarf in the style that is pinned on the back of your neck because it will emphasize the chin and the jaw. It is better to wear a visor *ciput* [a base layer with a built-in brim] with a wide visor that covers the ears so that it will give the impression that the face is longer and oval (Rahmadini and Arsminda 2011, 6; translation mine).

There are many options such as close-fitting or loose headscarves, pinned under the chin or not. Designers seasonally create new versions using distinct fabrics, patterns, and colors. Their goal, according to Iran Mutiara, designer of the popular Islamic brand Irna la Perla, is "to encourage all the Muslim girls around the world to cover up with the edgy kids of style."[6] Particularly in

[6] See the Ina's Scarf website, http://inasscarf.blogspot.com/2011_05_01_archive.html. Accessed June 12, 2011.

vogue in May 2011 were "ninja *ciputs*," tight base layers, worn under voluminous pashmina shawls and "bun *ciputs*," skull caps with stuffed buns that give the impression of a bun under the headscarf and are worn, according to Rania, "so our *hijab* will look more interesting. It is not interesting to look at someone with a flat head."

In addition to the importance of transitioning to veiling, the quest to find a flattering style, and the abundance of advice literature written in the vernacular, another distinct aspect of Indonesian fashion veiling is that the focal point of Islamic dress, a headscarf (or *kerudung* in Indonesian), is becoming a fashion trend in and of itself. Since the headscarf is so popular in Indonesia, some women, especially young girls, wear it merely as a fashion statement. As Rania puts it, "Women who wear *jilbab* can be divided into two groups: those who always wear *jilbab* and those who wear it occasionally, whether that is only for Eid [an Islamic holiday], to mosque, or just to go to school." During focus groups, women often spoke of girls who only occasionally wore *jilbab* as a fashion statement, and for whom the "look" of a headscarf, its color and pattern, its drape or volume, motivates their fashion decision. It might simply be that a particular outfit is "more fashionable" with a headscarf than without. In other words, in Indonesia prevalent clothing trends cast the Islamic veil as "in the mode," and thus the veil may be worn merely for its aesthetic value.

Lessons of the Hijabi Fashionista

Taken together, the way in which women have influenced Indonesian Islamic fashion (through their choice to dress modestly, their production of advice literature, their justification of flattering styles, and their establishment of headscarves as a trend) can be seen as one way women are making significant contributions to public moral norms. These lessons of *hijabi* fashionistas encourage us to rethink (1) the role of bodily practices in the cultivation of character, (2) the definition of modesty, (3) the relationship of beauty to virtue, and (4) the role of virtue in politics.

RETHINKING BODILY PRACTICES

Bodily practices are central to Islamic character formation. Take the scriptural basis for modesty in the Qur'anic verse 24:30–31: women are asked to *do* certain things in order to *be* a certain way. Many Indonesian women report that in the beginning the bodily practice of veiling is difficult. It takes a conscious effort. They feel awkward. They are unsure they are "doing it correctly." Veiling can feel unnatural at first if a woman is concerned with looking unattractive, for instance, indicating she has some measure of vanity. But after repetition, it is unveiling that feels unnatural, an indication that the dispositions associated

with veiling, such as modesty or shyness, have been successfully cultivated. Over time, a woman acquires a habit so that she wants to veil and unveiling is what feels uncomfortable. This marks a successful formation of character through bodily practice and encourages all virtue ethicists to think carefully about their own theoretical understanding of the relation of bodily practices and character formation.

The central role of bodily practices in the cultivation of Islamic modesty reminds us that religions are not concerned only with right beliefs, but also right action. Islam is no exception, and Muslims believe that completing certain bodily actions—such as prayer, fasting, alms-giving, and so on—are the "pillars" of a pious Muslim life. In Islamic ethics this is formulated along the following lines: belief, understanding, discussion, or persuasion are not enough to transform a person; repetitive behavior and physical habits are also part of moral development (Metcalf 1994, 710). The core idea is that practices modify and transform the person who does them. In this way the prostration of daily prayer cultivates humility and submission in the person who prays; fasting during Ramadan cultivates devotion; almsgiving cultivates generosity.

Often described in classic Islamic texts using the analogy of health and disease for virtue and vice, bodily practices are conceptualized as integral parts of character formation. Some virtues can only be formed through proper bodily practice just like some aspects of bodily health can only be acquired through exercise. Exercise that raises one's heart-rate strengthens the heart. Lifting weights builds bone density. Physical actions, while necessary, might not be sufficient for creating health, but they are part of a "healthy" lifestyle. In a similar way, the presumption of public veiling is that part of a virtuous lifestyle involves doing certain things, including dressing a certain way.[7] There is still a sense in which intent is valued, but intent does not necessarily need to precede physical actions as part of the cultivation of character. In this view, the idea that one might need to "fake it until you make it" is not ethical insincerity, but rather a natural part of the process of character formation.

This might not seem particularly innovative, since it follows closely Aristotle's view of the role of habituation in the moral life. But this emphasis on personal bodily actions over human intent and divine will does create a

[7] The tenth-century Islamic thinker Ibn Miskawayh (d. 1030) explains this as follows:

> Another obligation incumbent on the person who seeks to preserve the health of his soul is to apply himself to a duty relating to the theoretical part [of knowledge] as well as to the practical—a duty, which he should not, under any circumstances, be allowed to neglect, so that it may serve the soul as physical exercise is pursued to preserve the health of the body. Physicians ascribe great importance to exercise in the preservation of the health of the body, and the physician of the soul attributes even greater importance to it in the preservation of the health of the soul. (Ibn Miskawayh 2002, 159–160)

theological tension. For Aristotle, as for all Greek thinkers, humans have the ability to be virtuous because of our rational capacities. However, in the Islamic worldview good character is not only about intellectually discerning the right thing and acting on it but also, and more importantly, being in right relationship with God. Character formation for Muslims is a process of perfecting the soul for God, who is *the* perfection. Thus a view of virtue that requires bodily actions sets up a tension: is God or individual responsible for the cultivation of character?

RETHINKING MODESTY

When we speak about the modesty of Islamic veiling, we are really speaking of two separate virtues: (1) general modesty that looks like humility, entailing an underestimation (or correct estimation) of one's self-importance, for which the corresponding vice is vanity or overranking, and (2) sexual modesty that looks like chastity, entailing a properly demure behavior toward the opposite sex, for which the corresponding vice is promiscuity. From a feminist point of view, both senses of modesty are potentially problematic for women: (1) if modesty involves an underestimation of self-importance it can quickly morph into underestimating self-worth, which is particularly problematic within existing contexts of structural gender injustice (Young 2006); and (2) sexual modesty often assumes women's submission to men or at least that women's bodies are more sexually charged than men's.

But fashion veiling turns some of this on its head. For one, the women's practice and justification of fashionable or stylish veiling uncouples these two senses of modesty: the *hijabi* fashionista asserts her self-worth even as she attempts to dilute her public sexuality. This recasts fashion and style as self-care and reminds us that being beautiful is not the same thing as being promiscuous.

Second, submission to men is not part of the Indonesian veiling fashion conversation. For every woman I interviewed, the purpose of stylish pious dress was to gain the respect and admiration of her Muslim female peers. This helps counter the impression that modest dress is always and everywhere imposed on women by the men in their lives.

Third, a cross-cultural approach to veiling as one means to the cultivation of good character highlights local interpretations of modesty. Most people associate veiling with modesty, but what modesty is, how it is demonstrated, and how it is judged differs greatly among various cultural contexts. This is in part because of the range of interpretations of the Qur'an's directive to veil. Whether or not there is a duty to veil, and what such a duty might entail, depend on how a community or individual interprets which aspects of a woman's beauty must be covered. In other words, modesty is universally acknowledged by Muslims as an important virtue for a good character; however, what

virtue entails and how it is habituated are culturally specific, influenced by politics as well as the visual culture.

In Indonesia, this Qur'anic reference to covering is interpreted as a directive to literally "cover with cloth." Most Indonesians do not think it is necessary to cover a woman's shape. As Rania, the Indonesian fashion blogger, puts it, "Islam has no specific requirement for *hijab*, just that we should cover." Tight *mansets* (which are basically thin, tight, long-sleeved t-shirts), belts, and even leggings are not only common; they are considered pious. That a woman would and should find a fashion style that makes her more beautiful, that even flatters her figure, is accepted and even expected because fashion is made permissible, even desirable, by logically disconnecting "attractiveness" from "sexual promiscuity."

This was particularly striking to me given my research on veiling in other contexts. In Iran, for example, where public veiling is mandatory, local norms of modesty are connected to hiding a woman's shape. Iranian authorities are concerned with any sort of Islamic dress that is tight. Belts are not allowed. A tight *manset* or leggings would be read as improper, at best, and vulgar, at worst. While these styles might physically cover a women's body, they accentuate a woman's form. In Indonesia, however, this same clothing is worn as part of *jilbab*. In these ways, thinking about the diversity of Muslim fashion allows us to see that modesty and decency may be universal norms, but their content is context specific.

In terms of virtue, this has significant implications. For one, it shows how virtues can be universally acknowledged and yet culturally influenced. By wearing flattering *jilbab*, and arguing it is pious to do so, Indonesian women are redefining what counts as modest behavior. No longer shyness or being reserved in manner, modesty becomes proper bodily maintenance, much like an exercise routine, that helps maintain a good character.

RETHINKING BEAUTY

Part of the reason veiling that flatters is accepted in Indonesia is because women insist, both through their sartorial practices and in their discussion of *hijab*, that beauty means something more than "superficial" appearance. An Indonesian book published in 2011, whose title can be translated as *60 Common Veiling Mistakes*, explicitly makes this argument when it discusses two sources of a woman's attractiveness and beauty, "namely the physical beauty and inner beauty, or the beauty that comes from the inside." Outer beauty, the authors tell us, "can be seen in the face, manner of dress, and the body," while inner beauty "can be seen from the way she behaves, talks, and also her polite, soft, and proper words" (Fitri and Khasanah 2011, vii; my translation). It is clear that physical beauty alone is not the goal for a good Muslim woman; rather, it is cultivating a deeper inner beauty that is connected to her virtue.

But it is also clear that the two are not unrelated in the advice literature. As another publication declares, "when that inner beauty is surrounded by beautiful designs, it can radiate even more" (translated by Jones 2010, 105). In other words, outer beauty can intensify inner beauty.[8]

Fashion veiling is not a litmus test for morality, just as someone who is outwardly beautiful can be inwardly vicious. But at least some Indonesian women argue that physical beauty "realigns image and substance in a morally true order" (Jones 2010, 106), so that there is no conflict between being and looking good. In other words, although beauty is not itself a virtue, it is being conceptualized as an outward sign of a character trait or disposition. Similar to an Aristotelian conception of art as something that can express normative ideas (*Poetics* 25), some of the women I study see beauty work as an art form aimed at understanding God's creation, or stated otherwise, as a visual representation that can invoke enlightenment in audiences about complicated moral ideas.

The concept of beauty does substantial theological work, even if this work is unintentional. For instance, the linking of outer and inner beauty connects ethics and aesthetics in an intimate way, so that Islamic dress can be understood as pursuing both ends. Thus fashion-veiling might be understood as contributing a type of feminist virtue ethics that recuperates a rich Islamic tradition of aesthetic sensibility and performance for feminist ends. For example, greater attention to aesthetics could allow for moral theory and moral epistemology to be grounded in part on the human capability to appreciate aesthetics instead of merely our rational capabilities. And this would be a very radical thing indeed, at least in relation to a Kantian ethical theory.

RETHINKING POLITICAL

Fashion-veiling illustrates how visual bodily practices that cultivate character can also express normative tensions and ambiguities. Bourdieu wrote that visual culture could be a realm of "permanent revolution" (Bourdieu 1996, 239) in which moral agents use innovations in style and genre to oppose existing aesthetic conventions and thus challenge doctrine, authority, or specific

[8] Interestingly, this is not an idea that is foreign to Islam, and can be seen as the recuperation of some aspects of the twelfth- century Islamic thinker, al-Ghazali's (d.1111) concern with aesthetics. Take the following excerpt from his major work, *Revival of Religious Sciences*: "The interior luminescence illuminated the exterior by giving it ornamentation and luster . . . if a person has no humility in his heart do not expect to find it in his limbs. And whose heart is not a niche for divine lights, the beauty of prophetic example will not radiate" (quoted in Moosa 2005, 221). This connection between inner virtue and outer "luminescence" is grounded for al-Ghazali in what he calls *dhawq*, a transcendent aesthetic sensibility that Ghazali prefers over reason as the grounding of autonomy. Another way of saying this is that in contrast to Kant, for instance, al-Ghazali makes aesthetic appreciation versus simply rational capacity central to the human condition. The purpose of aesthetics for al-Ghazali is to help us access intuitive knowledge which we cannot come by through other discursive means such as Islamic jurisprudence (Moosa 2005, 235).

moral teachings.[9] Another way of putting this is that the performance of something in "bad taste" becomes a critique of the moral authority of the elite. Veiling fashion makes these sorts of political statements. As a form of "creative conformity" (Bucar 2010) or "accommodating protest" (Macleod 1991), veiling successfully "rebels" only if the woman first has mastery of the local presumptions about the reason and power of veiling. This means fashion lies between consensus and rebellion, as a medium of tactical critique that still plays by existing visual and ethical rules. It may thus prove to be subtle and thus more successful than other forms of feminist politics.

It is not possible to make universal claims about the ethical critique of fashion-veiling because what the Islamic veil means in a particular social context reflects local politics, history, and culture. We can, however, consider the power of these sartorial practices within a given context. According to Rania, Indonesia *jilbab* fashion is an explicit critique of anglo conceptions of beauty. She argues, "*[J]ilbab* attracts younger generations who want to protest the image of anglo beauty in the media and TV. Wearing *hijab* we are free." For Rania this critique is not only viable within Indonesia, which is demonstrated by the fact that she posts in English "to help *hijabis* all over the world."

In contemporary Indonesia fashion veiling and its justification are also implicit critiques of the moral authority of the male elite. Women are working to decide the virtue and aesthetic meanings of flattering Islamic dress. This occurs within a web of norms conveyed through marketing campaigns, religious instruction, and popular culture, but the ultimate synthesizer of these norms is not the scholarly male elite but, rather, the Indonesian woman who decides how to veil.

This has some intriguing implications for ethics. For one, it shows how everyday practices can have political implications. Second, fashion veiling shows how the cultivation of character is not focused only on self-formation. Dress is about aspiring to a mode of public presentation, not merely the display of personal moral norms or virtues. Character is formed in part by training that occurs in public, with others, and ideally results in a form of corporate virtue.

CONCLUSION: FEMINIST VIRTUE ETHICS MOVING FORWARD

In this chapter I have used virtue ethics to explain the popularity of women's fashionveiling in Indonesia. I also suggested that *hijabi* fashionistas are making significant contributions to public moral norms that encourage us to rethink

[9] Bourdieu developed this line of argument by adopting Max Weber's distinction between priests and prophets: avant-garde artists are the prophets who rebel against the established authority of the priests. Both priests and prophets share an investment in the visual field, the *illusio* (Bourdieu 1996, 227).

bodily practices, modesty, beauty, and politics in the cultivation of character. My hope is that this work can contribute to a growing understanding of the role of gender in virtue ethics more broadly.

Many understandings of virtue developed out of theories that assumed human flourishing was something only open in its fullest form to men, based on biological and metaphysical claims about gender.[10] This is no longer the dominant presumption in virtue ethics, but there are some important holdovers, such as gendered virtues that are argued to be differently important to women and men.[11] For instance sexual modesty is a virtue for both Muslim men and women, but the tradition has spent far more time debating what is required for women to cultivate this virtue than men. There remains important work to be done in articulating how or why this might be based on outdated notions of women's bodies as fundamentally more sexually potent and of men as having very limited self-constraint.[12]

References

Ali, A.Yusuf, trans. 2001. *The Holy Qur'an*. Hertfordshire: Wordsworth.

Annas, Julia. 1993. "Women and the Quality of Life: Two Norms or One?" In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen, 279–296. Oxford: Clarendon Press.

Bourdieu, Pierre. 1996. *The Rules of Art: Genesis and Structure of the Literary Field*. Translated by Susan Emanuel. Stanford, CA: Stanford University Press. Originally published in France by Editions du Seuill as *Les Règles de l'art* in 1992

Bucar, Elizabeth. 2010. "Dianomy: Religious Women's Moral Agency as Creative Conformity." *Journal of American Academy of Religions* 78 (September): 662–686.

Bucar, Elizabeth. 2012. *The Islamic Veil: A Beginner's Guide*. Oxford: Oneworld.

---

[10] Take the ancient Greeks, who considered slaves, women, and children as morally inferior to male citizens and property owners. In the *Generation of Animals* Aristotle calls a woman a deformed infertile male, and in *History of Animals* he asserts a man represents the perfection of human qualities and capacities. See Warren (1980, 38–39).

[11] This is not just a challenge for religious communities; in fact, the philosopher Julia Annas asserts that every society has "two actual norms for human life," one for men and one for women (Annas 1993, 279). These two norms are not necessarily unjust, but as ethicists, we will want to pay close attention to what ideals of human nature they assume and how they affect the quality of individual lives.

[12] This chapter was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. I would like to also thank the editors of this volume for their thoughtful comments on earlier drafts. I presented earlier versions of this essay in (1) Northeastern Humanities Center's Faculty works in progress series, (2) "Re-Orienting the Veil," a conference co-sponsored by Duke University and University of North Carolina at Chapel Hill, (3) the 2014 Society for the Study of Christian Ethics' Annual Meeting in Seattle, and (4) the 2013 Character Project Conference at Wake Forest University. I thank the organizers for the invitations to present and the participants for their insightful feedback.

Fitri, Idatul, and Nurul Khasanah. 2011. *60 Kesalahan Dalam Berjilbab* [60 common veiling mistakes]. Jakarta: Basmallah.

Jones, Carla. 2010. "Images of Desire: Creating Virtue and Value in an Indonesian Islamic Lifestyle Magazine." *Journal of Middle East Women's Studies* 6 (3): 91–117.

Jones, Carla. 2007. "Fashion and Faith in Urban Indonesia." *Fashion Theory* 11 (2/3): 211–232.

MacLoed, Arlene Elowe. 1991. *Accommodating Protest: Working Women, the New Veiling, and Change in Cairo*. New York: Columbia University Press.

Mahmood, Saba. *Politics of Piety: The Islamic Revival and the Feminist Subject*. Princeton, NJ: Princeton University Press. 2005.

Metcalf, Barbara. 1994. "'Remaking Ourselves': Islamic Self-Fashioning in a Global Movement of Spiritual Renewal." In *Accounting for Fundamentalisms*, edited by Martin Marty and Scott Appleby, 706–25. Chicago: University of Chicago Press.

[Ibn] Miskawayh, Ahmad ibn Muhammad. 2002. *Tahdhib al-akhlaq* [The refinement of character]. Translated by Constantine Zurayk. Chicago: Kazi.

Moosa, Ebrahim. 2005. *Ghazali and the Poetics of Imagination*. Chapel Hill: University of North Carolina Press.

Rahmadini, Kurniati, and Arsminda, Tirza. 2011. *Aneka Kreasi Kerudung Cantik: Gaya dasar memakai kerudung segi empat & selendang*. Jakarta: Demedia Pustaka.

Scott, Joan. 2007. *Politics of the Veil*. Princeton, NJ: Princeton University Press.

Tarlo, Emma. 2010. *Visibly Muslim: Fashion, Politics, Faith*. Oxford: Berg.

Warren, Mary Anne. 1980. *The Nature of Woman: An Encyclopedia and Guide to the Literature*. Inverness, CA: Edgepress.

Young, Iris. 2006. *Responsibility for Justice*. Oxford: Oxford University Press.

PART } VIII

Challenges to Character and Virtue from Neuroscience and Situationism

28 }

Character Traits and the Neuroscience of Social Behavior

Daniel J. McKaughan

Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant, in some cases using custom as a test, in others perceiving them from their utility. It is the same thing which makes us mad or delirious, inspires us with dread and fear, whether by night or by day, brings sleeplessness, inopportune mistakes, aimless anxieties, absent-mindedness, and acts that are contrary to habit.

—HIPPOCRATES, *FIFTH CENTURY BCE*

1. Introduction

Serious interest in integrating virtue theory with an empirically informed moral psychology runs all the way back to Aristotle and has been a critical point of emphasis in its twentieth century revival (Anscombe 1958). Discussion of the situationist challenge to virtue ethics, led by John Doris and Gilbert Harman, has brought questions about the bearing of experimental results in social psychology into sharp focus in the literature on virtue theory (Doris 1998, 2002, 2010; Harman 1999, 2000, 2001, 2003, 2009). The person-situation debate is now informed by a vast body of descriptive empirical research and, as matters now stand, recognition of the reality and importance of personality traits in this once hotly contested area continues to inform the research and thinking of many psychologists (Fleeson 2001; Fleeson and Noftle 2008; Fleeson and Gallagher 2009). Despite some deep concessions about the nature, robustness, and prevalence of character traits, the trend among virtue theorists is to construe character traits as personality traits of a particular sort—traits

the acquisition and maintenance of which is such that makes them appropriate targets of normative evaluation. One feature of the best current defenses of the psychological reality of character traits (e.g., Russell 2009; Snow 2010; Miller 2013; Miller 2014), is that they bottom out in folk-psychological mental dispositions—as do the kinds of psychological theories on which they draw (e.g., the cognitive-affective personality system, or "CAPS," model advanced by Mischel and Shoda 1995). However, these developments leave open the question of how personality traits (and whatever mental dispositions that might underlie them) will turn out to relate to a detailed biological understanding of their causes.

Presumably, there is some story to be told about how character traits—like all personality traits—get instantiated biologically. As Colin DeYoung, a proponent of the emerging field of personality neuroscience notes, "A causal theory of personality could be based on psychological constructs alone, but this would ignore the fact that the brain is the proximal cause of behavior and experience" (DeYoung 2010, 1176). But what significance might this have for ethics? Could there come a time, as Owen Flanagan put it to me, when instead of saying things like "That's just what I expected of Charity—she is so generous and trusting," we would comment instead on her high endogenous oxytocin levels or remark, "What a magnificent distribution of oxytocin receptors in her nucleus accumbens"? Or are such accounts, for issues of interest to moral philosophers, just so much irrelevant detail? My bet is that in the years ahead, discussions about the long-term viability of virtue theory, including questions about whether character traits can continue to play a central role in explaining and predicting human behavior and whether virtue theory can re-stake its claim both to closely mirror ordinary practice and to be aligned with a realistic and compelling moral psychology, will increasingly focus on what we are learning about the brain.

In the present chapter, I attempt to set the stage for such discussions by examining some recent claims about the significance of research on the neuropeptide oxytocin for ethics (section 2). Section 3 explores how different ways of understanding the project of virtue theory affects questions about the relevance of empirical findings. The assessment I offer pulls in two directions. On the one hand, in section 4, I identify a set of issues in which neurobiological accounts of the proximate causes of social behavior could, in principle, be relevant to virtue theory.[1] On the other hand, I resist buying into some of the

[1]For reasons of space, the present chapter focuses on proximate causes of social emotion, cognition, and behavior. But there are also important questions about the significance of relatively distal causes, for example, of oxytocin in the evolution of social attachment and moral sensibilities for moral philosophy (Churchland 2011a; McKaughan 2012). Although, in my view, we face moral and evaluative decisions, including judgments about who we should aspire to become, that cannot simply be read off of scientific descriptions of humans as we are, there are nevertheless a number of ways in which such information can be relevant to ethics. As Hume taught us, you cannot derive an ought from an is.

hype that has surrounded recent claims about the significance of neuroscience, arguing that, so far, a lot of the results of these studies are open to interpretations that are not nearly as revisionary for our commonsense understanding of persons or rational action as some would have us believe. Section 5 directly considers recent empirical work that establishes some important associations between oxytocin levels and the behaviors used to measure trust and generosity in humans. I argue that cases of exogenous administration of oxytocin can be used to raise a challenge to traditional virtue theory that is already well known in the situationist debates and that can be addressed with the same kind of reply. In particular, Christian Miller's (2013, 2014) Mixed Trait framework stands up quite well when applied to an area of research that he has not yet considered.

2. The Moral Molecule? Claims about the Profound Significance of Oxytocin

Let's begin by considering some claims about the profound significance for ethics of oxytocin, a peptide hormone that plays a role in chemical signaling in the brain. Oxytocin plays a variety of roles in human physiology, both while circulating in the peripheral bloodstream and as a chemical signal in the brain. Oxytocin is released in response to sexual stimulation, when the uterus dilates to give birth, and during nursing. But oxytocin is also involved in relieving anxiety and in complex social behaviors, such as maternal nurturing. Recent attention to oxytocin stems primarily from its role in mammalian social attachment, such as pair bonding and parental care, and for success that scientists have started to have in elucidating the mechanisms in neurobiological accounts of the proximate causes of complex social behaviors like these. Owing largely to experimental work that establishes some important associations between oxytocin and human prosocial behaviors

---

Nevertheless, there are facts about what you and I value. Some of these concern matters that we care deeply about and endorse upon reflection and thus can provide us with reasons for (not just causes of) action. As Harry Frankfurt observes:

> Nearly everyone cares about staying alive, for instance, and about avoiding severe injury, disease, hunger, and various modes of psychic distress and disorder; they care about their children, about their livelihoods, and about how others think of them. Needless to say, they generally also care about many other things as well. (Frankfurt 2004, 27)

We do develop deep social attachments and many of our concerns go well beyond self-interest. Moreover, given that I care about something, there are often objectively better or worse (or more or less effective) ways of trying to achieve that goal. So, although facts about the evolution of social attachment can't tell us what *to* value, they can tell us about what we *do* value and, from there, a lot about how best to achieve those ends.

such as trust (Kosfeld et al. 2005; Zak 2005; Zak et al. 2005; Zak et al. 2006), generosity (Zak et al. 2007) and reciprocity (Zak 2011), oxytocin has been characterized as a "trust hormone," "cuddle chemical," "love drug," or even as "The Moral Molecule"—metaphors that seem to promise a panacea for many of our social ills. Attention-grabbing though they are, these beguiling catchphrases deserve scrutiny, since they can frame the discussion of larger issues of social and political interest, such as how we as a society should understand ourselves as persons or moral agents, in ways that become as much sources of confusion as illumination (McKaughan 2012; McKaughan and Elliott 2012, 2013).

Paul Zak's research group at Claremont Graduate University has been a central contributor to the investigation of the relations between oxytocin levels and behavior in humans, so let us begin by considering Zak's characterization of oxytocin as "the Moral Molecule." Discussing an envisioned "new science of what makes us good or evil" in *The Moral Molecule: The Source of Love and Prosperity,* Zak writes:

> My research had demonstrated that this chemical messenger [oxytocin] both in the brain and in the blood is, in fact, the key to moral behavior . . .. Am I actually saying that a single molecule—and, by the way, a chemical substance that scientists like me can manipulate in the lab—accounts for why some people give freely of themselves and others are coldhearted bastards, why some people cheat and steal and others you can trust with your life, why some husbands are more faithful than others, and, by the way, why women tend to be more generous—and nicer—than men? In a word, yes. (Zak 2012a, x–xi)

Here is how Zak described the significance of his work in a *Wall Street Journal* interview:

> This book is about my unlikely discovery of a brain chemical called oxytocin that motivates people to engage in moral behavior. So it sheds new light on the longest debate since humans have been having debates, which is: are we good or evil? And we show that when you turn on oxytocin in the brain people engage in virtuous or moral behaviors . . .. We can develop stimuli in the laboratory that cause the brain to release this chemical and we developed a nasal inhaler with which we can shoot it into the brain safely. And in both cases when oxytocin is raised people are more generous, they are more compassionate, and in particular they are empathic. That is, they connect better to people emotionally. And we treat others that we are connected to emotionally, well . . .. Using the oxytocin nasal spray we can actually turn on moral behaviors like a garden hose. We can make people more trusting, more compassionate, more generous, more empathic. So we know there is a direct relationship

> between oxytocin in your brain and these particular moral behaviors. (Zak 2012b)

These are interpretive claims about the philosophical significance of a particular body of scientific research. The main claim is that oxytocin plays a central role in the causal explanation of a specific, though widely varied, set of "virtuous or moral" behaviors. Zak seems to be thinking about moral action, characterized in behavioral terms, primarily in terms of behaviors that benefit other people. But he also thinks that oxytocin makes a motivational difference, not so much by any association with the thought that what one is doing is morally right or good, but through some sort of causal influence on emotions or sentiments, such as empathy or compassion. The sense in which Zak thinks that oxytocin "accounts for why" people do or do not engage in these behaviors appears to rest less on an ability to actually state specific and detailed explanations (or even descriptions) of whatever complicated underlying neurophysiological mechanisms might together yield an understanding of why the behaviors occur, than on the claim to have identified oxytocin as a causal influence, the manipulation of which allegedly exerts a high degree of control over the behaviors in question.

For a number of reasons, the characterization of oxytocin as the Moral Molecule, with the definite article and capital *M*s, needs qualification. Whether one regards such characterizations as harmless and amusing or eschews their use, there is much to be said for recognizing the ways in which the unreflective use of such labels can mislead us.

First, when confronted with claims that genes or chemicals "control," "determine," or "account for" a given behavior—or about the capacity of neuroscientists to turn moral behavior on and off "like a garden hose"—we ought to think carefully about whether they correctly characterize the relative strength and scope of the causal influence in question. It should go without saying—and it would go without saying if we did not dub oxytocin the Moral Molecule—that oxytocin is far from being the only molecule, let alone the only causal factor, involved in producing moral feelings and behavior and that any effects oxytocin has are dependent on the presence of other factors.[2] In mammals oxytocin

[2] Notice that, although Zak tends to talk as though it is primarily the levels of oxytocin that make a causal difference, claims about the causal effects of a single molecule are context-dependent. The effects of oxytocin depend, for example, on the density and distribution of its receptors and its activity can be activated or antagonized by other molecules such as estrogen and testosterone. Some of its effects depend on interaction with other biological pathways such as the dopamine reward system. Some inputs are, of course, more significant than others. But any behavioral effects for which oxytocin is an important causal factor are exerted within the dynamic environmental milieu of the developing organism.

Discussions about the kinds of causal influence that oxytocin might have on various brain processes and its relation to higher level cognitive aspects of decision making, self-control, and voluntary behavior of importance for issues of moral agency should not overlook the extent to which the neurophysiology involved in the oxytocin and vasopressin systems, including gene expression, is responsive to epigenetic and environmental factors. While, as Zak (2012a) points out, physical contact such as

exerts a stronger and more evolutionarily basic influence on maternal nurturing than it does on moral behavior. But, as Patricia Churchland points out, despite the genuine causal role that oxytocin plays in maternal behavior, given that so much more is required for successful maternal behavior (and the fact that it is not clear that oxytocin is even a necessary condition for such behavior in humans), the idea that a single chemical plays a privileged causal role amid the myriad of other important factors offers only a superficial and potentially misleading understanding of the phenomena in question (Churchland 2011b).

Second, an apt label should not present a one-sided picture of the goods or evils that our brain chemistry might influence us to do. Oxytocin appears to be involved in increasing, not only trust, empathy, and cooperation, but also less socially desirable responses, such as envy, gloating (Schadenfreude), in-group favoritism, and ethnocentrism (Shamay-Tsoory et al. 2009; De Dreu et al. 2010; De Dreu et al. 2011a, 2011b). Given this evidence, as Molly Crockett points out, the same standards for labeling oxytocin a "moral molecule" would permit its characterization as an "immoral molecule" (Crockett 2012).

Third, there are still questions about whether the effects of oxytocin are specific for the aforementioned sorts of narrowly construed cognitive states and behaviors (e.g., trust) or could instead be explained by a more general effect on social motivation, anxiety, or arousal—a topic of on ongoing discussion in a recent special issue on oxytocin, vasopressin, and social behavior in *Hormones and Behavior* (Young and Flanagan-Cato 2012; Churchland and Winkielman 2012).

Fourth, a number of other pertinent questions about the effects of oxytocin remain in a state of flux. For example, after noting some of the small effect sizes, situation-dependence, and apparent inconsistencies in the current oxytocin literature, some scientists have even argued that the effects of social

---

giving and receiving a hug or a back massage can temporarily elevate oxytocin levels (perhaps for twenty minutes, perhaps for an hour) other kinds of environmental stimuli can induce long term changes. For example, in voles, a recent study in *Nature Neuroscience* reported "the first direct evidence for epigenetic regulation of pair-bonding" (Wang et al. 2013, 919). When the voles mate this induces a process (histone acetylation) which loosens tightly packed stretches of DNA and increases transcription of the oxytocin receptor genes (*OTR*) in the nucleus accumbens. This increase in the density and distribution of receptors is associated with partner preference formation and mate attachment. Here a social behavior, in this case mating, can lead to lasting changes at the neuroanatomical level that influence future social behaviors. In humans, there is evidence that childhood neglect can affect not only emotional development but that "the vasopressin and oxytocin neuropeptide systems, which are critical in the establishment of social bonds and the regulation of emotional behaviors, are affected by early social experience" (Fries et al. 2005). Meadow voles reared by prairie vole parents both receive more parental contact than from meadow vole foster parents and display higher levels of parental care in adulthood themselves (McGuire 1988). Early social experience can also alter gene expression and affects neuroendocrine development and adult behavior, such as stress response and maternal care, in rats. Female pups receiving high licking and grooming show increased expression of oxytocin receptors (OTR) and provide higher levels of maternal care themselves (Francis et al. 1999; Francis et al. 2002; see also Cushing and Kramer 2005).

context and individual persons are significant enough that it is a mistake to conceptualize oxytocin as having broad and invariant effects on human behavior (Bartz et al. 2011a; see also Bartz et al. 2011b). There is evidence to suggest that oxytocin actually hinders trust and cooperation in individuals with borderline personality disorder and that oxytocin "does not uniformly facilitate trust and pro-social behavior in humans; indeed, OXT [oxytocin] may impede trust and pro-social behavior depending on chronic interpersonal insecurities, and/or possible neurochemical differences in the OXT system" (Bartz et al. 2011a). Moreover, as Bales and her colleagues (2013, 180) point out, "Oxytocin (like other peptides) can produce opposing effects at different dosages," and its short-term and long-term effects might well differ substantially. For example, their research on prairie voles found that, while short-term oxytocin treatment increased affiliative behaviors with familiar partners, as seen in humans, in the long term, low doses actually reduced partner preference. The authors conclude that "[l]ong-term developmental treatment with OT [oxytocin] may show results different to those predicted by short-term studies, as well as significant sex differences and dosage effects" (Bales et al. 2013, 180).

Although many will balk at Zak's daring interpretive claims, there are a number of ways in which findings in neuroscience—such as that feelings of empathy can vary with, and be at least partially mediated by, oxytocin—might have significance for virtue theory. In the next several sections I offer a different assessment of how aspects of such work in neuroscience might bear on our understanding of character traits, returning specifically to the relevance of exogenously administrated oxytocin in section 5.

3. Skin in the Game? Virtue Theories and Empirical Content

The significance that advances in neuroscience might have for virtue theory depends on what one takes the project of virtue theory to be. Virtue theorists have often developed their theories in ways that have a significant stake in substantive claims from moral psychology—sometimes by making fairly specific assumptions about human nature.[3] There are, however, a lot of different ways in which virtue theories have been and can be developed. So of any particular theory, we should ask: What sort of empirical content does it have? What

[3] For example, far from endorsing an austere behaviorism, contemporary Aristotelian virtue theorists typically take internal mental states of the agent, such as motivations and intentions (or emotions, desires, beliefs), to play a role in in practical deliberation and to be crucial for moral evaluation of their actions and character (e.g., Adams 2006; Annas 1993, 2011; Hursthouse 1999; and Slote 1992). This presupposes a rich inner psychology with distinguishable mental states. The virtuous person is not simply someone who acts appropriately in a given situation; the continent (*enkratic*) person also (reluctantly) does that. The fully virtuous person thus acts for the right sorts of reasons and from the right sorts of motivations.

observations could, in principle, undermine or vindicate its key concepts or assumptions?

We can distinguish between two quite different goals or expectations one might have for a virtue theory. Consider, first, what we might call a *purely aspirational* project in virtue theory. It holds up exemplars for our imitation, perhaps drawn from real life or fiction, and invites us to share a rich range of value judgments about what traits of character (or acts) we should admire and cultivate or eschew. Inspiring portraits of moral excellence, or of weakness and deficiency, on display in heroes, villains, and ordinary folk might prompt us to reflect upon (and even motivate us to shape) our lives, how we treat each other, how we regulate our emotional responses, and our practical reasoning in quite particular ways or teach us to more readily identify the morally salient features of a situation. Such accounts will have something to say about the kinds of persons we should be striving to become, of the character traits we should aim to develop. Perhaps such projects will base an account of morally good or right motivation and action in a picture of the fully virtuous person. Certainly they could prompt us to reflect on what is really worthwhile or whether we have an adequate understanding of what flourishing consists in.

Clearly there is much of value in the aspirational project and most virtue theories will have aspects that fall under this heading. There would be ways of pursuing the aspirational project (for example, by making only normative claims or only quite vague or minimal assumptions about human nature) such that it would be relatively insulated from empirical challenge or refutation. Any stake that such a theory might have in claims about human nature might be so unspecific, non-committal, flexible, or peripheral that it would not much matter what is going on in our brains.[4]

But most virtue theories do not limit themselves in this way and instead seek to connect more deeply with the details of an empirically adequate moral psychology. As an example of one way in which contemporary virtue theories go beyond the purely aspirational project, consider the *explanatory project*. Some virtue theorists are invested in the task of developing a moral theory that will also serve to explain and predict behavior by appeal to dispositional traits. Here is what Alasdair MacIntyre takes his own project to be committed to in *After Virtue:*

[4] Even for ethicists seeking to limit themselves fairly strictly to the aspirational project, empirical information would be relevant to deciding how best to achieve our goals, for example, by identifying factors that might promote or hinder the process of becoming virtuous or of getting from point A to point B in the moral life. Moreover, if they accept certain other constraints on theorizing, such as ought implies can or that normative requirements should be psychologically realistic for creatures like us, the details about humans as we are might constrain the range of visions of who we could plausibly aspire to be. (Though, even portraits that present us with unattainable ideals might nevertheless move us to become at least a little bit better than we currently are.)

> [M]y account is Aristotelian in that it links evaluation and explanation in a characteristically Aristotelian way. From an Aristotelian standpoint to identify certain actions as manifesting or failing to manifest a virtue or virtues is never only to evaluate; it is also to take the first step towards explaining why those actions rather than some others were performed. (MacIntyre 2007, 199)

MacIntyre is confident that there is no replacing explanations of this kind: "Indeed without allusion to the place that justice and injustice, courage and cowardice play in human life very little will be genuinely explicable" and he goes so far as to say that explanatory projects in the modern social sciences which omit reference to them are, for that reason, "bound to fail" (MacIntyre 2007, 199).

Explanation and prediction (alongside understanding, evaluation, and imitation) are among the functional roles that Christian Miller envisions for character traits (Miller 2013, 12–13). So character traits are called upon to play a *causal* role in explaining and predicting some morally significant human behavior. Notice that this sort of project sees the task of ethical theory as part of a project that goes well beyond a concern with normative notions, such as the right and the good—with values rather than facts, with what ought to be rather than what is, and with what is desirable rather than merely desired. Key theoretical concepts, in this case a particular understanding of character traits, are assumed to correctly identify the true causes involved in a descriptive characterization of at least some of our behavior.

But what are character traits, more precisely? What sort of claim are we making about people when we attribute character traits to them or attempt to explain their behavior in those terms? One feature of the best current defenses of the psychological reality of character traits (e.g., Russell 2009; Snow 2010; Miller 2013, 2014), is that they cash out talk of character traits in folk-psychological concepts. Christian Miller, for example, takes traits to be constituted by, and perhaps identical with, certain mental state dispositions—clusters of dispositional beliefs and desires:

> [W]e can understand a character trait had by a person as roughly: A disposition to form beliefs and/or desires of a certain sort and (in many cases) to act in a certain way, when in conditions relevant to that disposition. (Miller 2013, 6)

Miller endorses a view of character traits according to which they are both real and causally active clusters of interrelated mental state dispositions, understood as relatively enduring and "stable mental attributes or structures had by a person which, when activated, causally give rise to relevant thoughts and (in many cases) behavior" (Miller 2014, 122). The causal powers that Miller ascribes to character traits are "grounded in mental structures in a person's mind" (Miller 2014, 126), and of course he leaves open the possibility, as he

should, that there are biological structures and processes that underlie and generate the mental ones.

These are ambitious goals, and the stakes are high. It would seem that explanatory projects of this sort will eventually need to integrate whatever claims they make about character traits with a biological understanding of the processes that underlie them. An allegedly more satisfying relation to the details of moral psychology has often been touted as a key advantage that virtue theory enjoys over its deontological and utilitarian rivals—an advantage that John Doris (1998, 520) takes to have been lost as a casualty of the situationist debates. A virtue theory that succeeds in offering, or in richly integrating itself with scientific theories that offer, psychologically realistic accounts of the role of motivation and deliberation in moral behavior or in offering substantive guidance about how to inculcate ourselves with the kinds of habits and dispositions that regularly lead to virtuous behavior would go a long way to reclaiming the alleged competitive advantage. An approach that could do this while at the same time closely mirroring ordinary moral intuitions, practice, and ways of understanding persons would have wide appeal. The danger here, of course, is that ethicists will get things wrong—and that they might wed themselves to assumptions that will come into competition with scientific theories of human behavior covering the same phenomenal domain.

4. Character Traits and the Potential Relevance of Proximate Causes of Complex Social Behaviors

Clearly there is some story to be told, as John Doris colorfully put it to me, about "what is going on in the meat" between our ears when we engage in moral behavior and reflection. But what significance could this have for virtue theory? There are a number of ways that neuroscientific accounts of the proximate causes of social behavior could, in principle, have revisionary implications for the explanatory project. Even while many empirical questions remain quite open, we can begin to map the conceptual significance of various radical scenarios for Miller's explanatory project.

Consider the ontological grounding of character traits. Miller takes traits to be constituted by certain clusters of interrelated mental dispositions. It seems clear that nothing in that view, just as such, would be incompatible with token-token identity (such that every character trait—or whatever cluster of mental dispositions that constitute the trait—is identical to some token set of neurophysiological states and processes). Suppose that the dispositional beliefs and desires that constitute a certain trait turn out to be reducible to brain states that are part of the physical system that oxytocin is acting on. Those brain states would then be the character trait. But that wouldn't be a

problem for Miller's view because they would be identical to the cluster of beliefs and desires that he has identified as the trait.

It might be that ascriptions of the attitude-types of folk psychology will map cleanly onto underlying brain states identified by neuroscience. But it also seems possible that things won't go so smoothly. Neuroscience might come to offer alternative explanations that replace those offered by the virtue theorists of our day or attempt to recover the phenomena we are talking about in disorientingly unfamiliar terms—such that personality traits or folk mental dispositions might not have any role in a complete neural level explanation of the mechanisms. Eliminativists, for example, claim that there are no such things as beliefs and desires, often on the basis of controversial predictions about how certain research programs will fare in the years ahead (Rorty 1965; P. M. Churchland 1981; P. S. Churchland 1986; Stich 1983). Here, the idea is that the folk belief-desire psychology we have inherited is a radically misguided theoretical framework for understanding human behavior, in that the mental states posited in ordinary discourse do not in fact exist. This framework, eliminativists predict, is destined not for reduction but for replacement by concepts arising out of a new and more empirically adequate science of the brain. Unlike token-token reductionists, eliminativists expect that the mental state terms of folk psychology will eventually be discarded—not because they have been reduced to brain states but because they no more refer to any real phenomena than does talk of witches or phlogiston. Obviously, if eliminativism with respect to the occurrent or dispositional mental states postulated by folk psychology turned out to be correct that would be a problem for an explanatory project of the sort Miller defends. Similarly, if epiphenomenalism were true, such that there are beliefs and desires along for the ride but these are explanatorily impotent and have no causal role in behavior, this would undermine Miller's project (and a lot of others).

With respect to the specters of eliminativism and epiphenomenalism, virtue theorists who, like Miller and MacIntyre, engage in the explanatory project in a way that throws their lot in with folk psychology can respond by making a strong and entirely sensible case for the reality and casual efficacy of dispositional beliefs and desires. Here one can point not just to the success of folk psychology at the everyday level but also to empirical support for psychological theories, such as the CAPS model proposed by Walter Mischel and Yuichi Shoda (1995), and to entire areas of psychology that take appeals to such items (e.g., cognitive-affective units) as beliefs, desires, emotions, and intentions to be not only legitimate and empirically well-grounded but also indispensable to understanding human behavior.[5] Folk psychology outlived a series

[5] Miller argues that "*using technical language, the CAPS model re-describes and finds supporting evidence for basic platitudes of commonsense folk psychology*" (Miller 2014, 108, 126; italics in Miller in both places. See also Johnson 1999, 449–450).

of behaviorist challenges in the twentieth century. Even after all these years, when it comes to explaining and predicting the vast majority of moral behaviors that ethicists care about, folk psychology arguably remains the best game in town. That seems like a defensible way to go. It would be nice, eventually, to have some sort of account of how the mental states relate to the neurophysiological states or, perhaps, a multilevel perspective on explanation that helps us to understand clearly how both kinds of factors, feelings of empathy and oxytocin levels, can legitimately enter into causal explanations of behavior. But it is reasonable to expect that dispositional beliefs and desires have some sort of neurobiological grounding, one that isn't merely epiphenomenal.

Another possibility is what we might call *conceptual fragmentation*. I have in mind a situation in which we come to realize that the categories we regularly employ in some domain turn out to encompass such disparate phenomena and to be produced by such different causal mechanisms that grouping them in the usual way either does not allow for reliable explanation and prediction or is superseded by separating the phenomena into more fine-grained scientific categories that are superior on these scores. Paul Griffiths argues that this is precisely what has happened to the ordinary concept of "emotion" in the course of investigating its biological underpinnings: the psychological states and processes grouped under the ordinary folk category of emotion (which includes items ranging from relatively spontaneous and unreflective forms of fear, disgust, or a flash of anger to less reflexive or more cognitively involved responses such as guilt, shame, or anger that take the form of a bitterly held grudge and remain for years) are not sufficiently similar to allow a unified scientific treatment (Griffiths 1997, 2004). One can well imagine the general category of character trait dispositions undergoing similar fragmentation when explanations are pressed to the level of neurophysiological mechanisms or even fragmentation with respect to particular character trait types. Just as Griffiths expects that the best scientific explanations of particular types of emotions, such as anger, very likely will differ from occasion to occasion, it might turn out that the best explanation for the group of occurrent psychological and behavioral responses that virtue theorists attempt to explain by appeal to a "courageous disposition" would involve different, perhaps vastly different, underlying psychological and neurophysiological explanations from occasion to occasion.[6]

As an illustration of how, in principle, such fragmentation might occur, consider *phronesis*—practical or moral wisdom—a traditional character trait the development of which is part of the goal of moral education. While

[6] This sort of fragmentation and accompanying explanatory obsolescence need not entail ontological elimination: traditional use of the term "jade" turned out to partially refer to two different kinds of minerals, jadeite and nephrite, each of which has distinct properties that are now investigated and explained separately from a scientific perspective. So too, Griffiths claims, with emotions.

character traits are often the objects of moral evaluation, some Aristotelian virtue theorists have held that the exercise of moral judgment itself involves or expresses certain character traits. There is a great deal of contingency and variability in the situations in which we find ourselves, and it is not implausible to think that discerning which features of a situation are morally relevant and making practical judgments about their significance will often involve certain wisdom-related traits. For Aristotle, excellence with respect to a trait like *phronesis* or practical wisdom involves both a kind of discernment or understanding of morally salient aspects of a situation and an ability to evaluate or assess or deliberate about whether acting in a given way in such circumstances is right or appropriate or such that it promotes *eudaimonia* or flourishing (Urmson 1988, 81–82). Yet it seems at least possible that a biological understanding of the processes involved in such judgments could lead to fragmentation with respect to these sorts of character traits. Suppose that Jack judges accidental harms to be less blameworthy than you take to be normatively called for, whereas Jill takes them to be even more blameworthy than you do. Virtue theorists of various stripes might attempt to explain these patterns in judgment by appeal to deficiencies or excesses in character traits related to moral judgment—traits that might also make predictions about their normative judgments in other sorts of cases. Now suppose we come to understand that,

(1) the observed variation in the extent to which individuals mitigate blame for accidental harms measurably and predictably correlates with identifiable genetic or neuroanatomical differences;

and that,

(2) biological theories explain some subset of phenomena that the postulation of (or appeal to) a character trait was supposed to explain and make predictions that are different from and clearly superior to an appeal to character traits.

In such circumstances, virtue theoretic and biological explanations could come into conflict in a way that leads to fragmentation and in which the former would end up looking a bit old-fashioned.[7]

---

[7] My point here concerns the in principle relevance of neuroscience for the predictive and explanatory value of character traits, including traits relevant to moral judgment. But, in case it helps some readers to see fragmentation as a live possibility, it should not go unnoticed that we already have evidence for (1).

Differences in the extent to which individuals mitigate blame for accidental harms are known to be associated with variation in the activity level and functionality of the right temporoparietal junction (RTPJ)—a part of the brain above and behind the right ear—that is normally recruited during certain kinds of mature moral judgments (Young et al. 2007; Young and Saxe 2008; Young et al. 2010). A recent study by Liane Young and her colleagues found that individuals with high functioning autism (ASD) judged accidental harms to be more blameworthy than 'neurotypical' individuals from the general population (Moran et al. 2011), while individuals diagnosed with psychopathy judge accidental harms

Assessing the potential implications of fragmentation for virtue theory is more complicated than for extreme outcomes such as eliminitivism and epiphenomenalism. On the one hand, there would be certain types of fragmentation with which Miller and others could live quite comfortably. Character traits are not just arbitrary groupings, and it would be surprising if fragmentation were to render such categories useless for addressing questions of interest to moral philosophers. For comparison, most of us expect there to be a terribly complex (and probably multiply realizable) neurophysiological story to be told about the dispositional habits and skills involved in knowing how to ride a bicycle or to do arithmetic (Flanagan 2009, 53). Yet, even if in some sense these categories did not cut nature at its joints, given our interest in riding bikes and doing math, it is hard to see how the gory details of neuroscience could render our ordinary use of them obsolete. On the other hand, we should recognize the possibility that traits might fragment in such a way that, owing to the availability of more specific predictive and explanatory categories, it begins to no longer seem useful to appeal to character traits. We can expect to learn a great deal more in the years ahead about how naturally occurring biological variation in brain structure and functionality affect not just moral behavior but also moral motivation and cognition. Discerning whether fragmentation is taking

---

to be less blameworthy than controls (Young et al. 2012). They suggest that the atypical responses of psychopaths may reflect characteristic deficits in empathic concern (so that accidental harms are more readily forgiven than in controls) and that deficits in perceiving the mental states of others characteristic of autism might explain why their judgments of accidental harm appear to be relatively more outcome-focused than controls (Gray et al. 2012b, 212; Gray et al. 2012a).

Because oxytocin is known to play a role in social perspective-taking and emotional perception, another recent study (Walter et al. 2012) adapted the experimental design of Young and Saxe (2009) to investigate potential associations between genetic variation in the oxytocin receptor gene (*OXTR*) and judgments about blameworthiness in cases of accidentally committed harm. One of the variants they examined is a single nucleotide polymorphism (SNP) in the promoter region of the oxytocin receptor gene (*OXTR*) named rs2268498. Some people have a cytosine base (C) in the DNA molecule at this position (chromosome 3p25) and other people have thymine (T) there. In this study, 154 participants were asked to make moral evaluations (e.g., assign blame) in response to vignettes that portray intentional harms and accidental harms. Walter and his colleagues found that participants with a particular genetic variant—people with a cytosine at this position on one or both of their third chromosomes—tended to rate accidental harm as significantly more blameworthy than individuals with a thymine at both these locations (Walter et al. 2012, 271). The results are statistically significant ($p = .001$): one would expect to observe such between group differences in mean ratings by chance, were there no relationship between this genotype and the behavioral phenotype, only about one in a thousand times. As the authors conclude, "individual differences in moral judgments that require the consideration of an agent's mental state" is partially explained by genetic variation in the oxytocin receptor gene (*OXTR*) and they take their findings as "evidence for a genetic contribution to moral judgments" (Walter et al. 2012, 271). Another recent study indicates that this same functional SNP in the OXTR promoter region (rs2268498), in interaction with 5-HTTLPR (a polymorphism in the promoter region of the serotonin transporter gene), influences negative affect as reflected in personality measures such as fear and sadness (Montag et al. 2011). Yet another study of rs2268498 found a relationship between oxytocin receptor genotype and recognition of facial emotion such that "carriers of the T-allele exhibited more accurate recognition skills than subjects carrying the CC-genotype" (Melchers et al. 2013).

place and whether there are any devils in the details of any actual and available science is a long-term project.

Advances in neuroscience might also be relevant to questions about responsibility for trait possession. How might empirical information, in combination with particular commitments of a given normative theory, lead us to reclassify the moral status of certain thoughts, motivations, and behaviors? Miller attempts, tentatively, to distinguish between character traits and personality traits more generally by some sort of criterion for normative responsibility or reasons-responsiveness (Miller 2014, chap. 1, sec. 1.2). There is disagreement about the extent and nature of self-control one must have over the acquisition, maintenance, or expression of the trait in order for normative evaluation to be appropriate. But the basic idea is clear: whereas it would be inappropriate to hold infants or certain adults with severe mental handicaps to ordinary normative standards for truth-telling, we do assess most people with respect to their tendencies in this domain—they are, in this respect, appropriate targets for our reactive attitudes (Miller 2014, 13).

A deeper understanding of the biological sources of individual traits might, then, lead us to reclassify a trait that we had once regarded as a virtue or vice (e.g., Charity's generosity or Henry's cowardice) in such a way that we would no longer regard praise or blame for its possession as appropriate. Perhaps what we have always chalked up to Charity's generosity owes more to her naturally high endogenous oxytocin levels. Given Henry's active amygdala, it may be much harder for him to suppress the fight or flight response than for others whose genetic and neurophysiological makeup, even prior to training, endows them with a calm in the presence of danger that better fits the profile typical of Special Operations Forces personnel. One can well imagine that natural variation in levels of activity, development, or functionality (or other sorts of differences between people's brains) might turn out to have a significant influence on relevant cognitive, affective, and behavioral responses.

Although such considerations illustrate the potential relevance of neuroscience for normative theory, the point is not that current work has rendered moral theories in which character traits play a central role problematic. Certain types of findings might occasion reclassification of the moral status of particular cases or even begin to blur the distinction between character traits and personality traits (e.g., if it came to seem in paradigm cases of character trait attribution that we lacked, perhaps due to strong genetic determination, the sort of control required to distinguish them from typical cases of non-character traits). Yet the guidance some of these very moral theories could provide in helping us to reassess the moral status of particular traits in light of a more accurate and complete understanding of their biological underpinnings might well yield the intuitively correct results. We do think it is appropriate to take genetic factors into consideration and inappropriate to hold certain people with severe mental handicaps to ordinary evaluative norms of truth-telling.

While new scientific information in conjunction with certain widely shared intuitions might lead us to rethink the moral status of some particular cases, it might in fact be quite hard for neuroscience to establish a level of causal determination over the acquisition, maintenance, or expression of a trait so sufficient as to make normative evaluation entirely uncalled for.

5. Oxytocin and the Situationist Challenge to Virtue Theory

Returning now to the experimental research undertaken by Zak and his colleagues, let us briefly consider one fairly clear way in which this research arguably bears on issues of interest to virtue theory. A number of recent studies have investigated the effects of oxytocin using intranasally administered oxytocin. Subject behaviors are then observed, and compared to a control group receiving a placebo, as they participate in various economic decision games. For example, oxytocin enhances interpersonal trust in a "trust game" in which subjects can choose to allocate money to a stranger, anticipating that their gesture might be rewarded with higher returns (Kosfeld et. al. 2005). Kosfeld and his colleagues report that subjects receiving oxytocin transferred money more often (45% of the time compared to 21% in controls receiving the placebo) and more of it—the average transfer of subjects in the oxytocin group is 17% higher than those of controls (Kosfeld et al. 2005). Moreover, 45% of subjects in the oxytocin group transferred the maximum amount, compared to just 21% in the placebo group. A subsequent study by Baumgartner et al. (2008) found that "subjects in the oxytocin group show no change in their trusting behavior after they learned that their trust had been breached several times while subjects receiving placebo decrease their trust" (Baumgartner et al. 2008, 639). In another kind of game, the "ultimatum game," participants presented with a sum of money are asked to propose how to split it with another stranger, knowing that the stranger has been asked to decide on a minimum threshold below which offers will be rejected and that neither participant will get to keep any money. Zak's research group reports that subjects who receive intranasal oxytocin are 80% more generous in ultimatum games than those given a placebo (Zak et al. 2007; see also Zak 2011).

One way in which studies demonstrating that external administration of oxytocin via a nasal inhaler can measurably increase generous behavior might pose a challenge to virtue theory is parallel to a challenge now familiar in the situationist debate literature. A number of studies by social psychologists have found that pleasant smells increase helping behavior (Guéguen 2012). Perhaps the most well-known of these experiments, a 1997 study by Robert Baron, presented passersby at a shopping mall with an opportunity to help retrieve a dropped item or to make change for a dollar. People were more likely to help in locations near businesses from which pleasant fragrances were emanating

(e.g., bakeries such as Cinnabon and Mrs. Field's Cookies) because, it is widely thought, pleasant odors put people in a better mood (Baron and Thomley 1994) and, in turn, people are more likely to be helpful when in an elevated mood (Harris and Smith 1975; Levin and Isen 1975; Weyant 1978; Bizman et al. 1980; Job 1987).

Studies of this kind suggest that most of us are, to a disturbing extent, influenced by these seemingly trivial features of the situation, which do not provide morally relevant reasons for helping. Classical virtue theory requires that the virtuous person not only behave well or appropriately but to do so for the right reasons and with good or appropriate motivations (Annas 2006, 516; Audi 2001, 82; Winter 2011, 14; Urmson 1988, 28). Someone who possessed the virtue of compassion, for example, would tend to be sensitive to morally relevant reasons for helping and not swayed so significantly by factors such as the variables in Baron's study. Moreover, these influences are often exerted without the actor's conscious awareness and have required carefully controlled observation not to be overlooked by psychologists.

Whether or not classical virtue theory is committed to widespread attributions of virtue or vice is controversial and, at any rate, something that many virtue theorists deny.[8] But situationists such as Gilbert Harman and John Doris have drawn on a large and varied scientific literature to argue that most of us do not have the traditional virtues. Just as Baron's study suggests that explanatory appeals to traditional virtues leave out an important factor—the causal influence of fragrances on helping behavior—Zak's nasal spray studies suggest that explanatory appeals to traditional virtues overlook the causal influence of oxytocin levels on generous behavior. Since, for a lot of participants in these studies, their trait-related behaviors depend significantly on the presence or absence of "enhancers" (e.g., activation by the oxytocin spray) in a way that does not track relevant moral considerations, the traits involved in explaining the behaviors would not qualify as traditional Aristotelian virtues.[9] Christian Miller (2013, 2014) accepts this line of argumentation and takes studies of this kind to suggest that most people do not have the classical virtues or vices (and indeed, reads Aristotle as agreeing that virtues are rare). Quite generally, in order for a person's character trait to qualify as virtuous, the frequency of

[8] For discussion of the possible grounding of such a reading in Plato's *Republic* and Aristotle's *Nicomachean Ethics* (e.g., 1179 b 11–16) and extensive references to debate about this point in the contemporary literature, I refer readers to Miller 2014, chapter eight, section three, "The Rarity Response and the Real Challenge to Virtue Ethics from Psychology."

[9] Miller (2013) introduces "enhancers" and "inhibitors" as technical terms to refer to influences on a person's motivation to perform or not to perform certain actions in trait-relevant situations relative to, other things being equal, what the person's motivational states would be in the absence of the enhancers or inhibitors (see, for example, Miller 2013 chaps. 2 and 8, sec. 2; see also Miller 2014, chap. 2, sec. 3. A discussion of the precise features of traits that get called into question by the situationist challenge can be found in Miller 2014, chapter four.

behavior related to that virtue should not depend on the presence of certain non-moral or morally problematic enhancers or inhibitors in such a way that, if these factors were not present, the frequency of virtuous behavior would significantly decrease or increase in situations relevant to the expression of the virtue. Based on his extensive examination of the psychological literature, Miller (2013, 2014) concludes that the overall balance of experimental evidence justifies us in believing that most people do not possess the traditional virtues or vices.

But Miller offers an alternative view, the Mixed Trait approach, which is capable of handling fragrance studies such as Baron's.[10] Unlike Harman and Doris, Miller (2013, 2014) does think that we have character traits, which are global in the sense that they exhibit both cross-situational consistency (e.g., trait-relevant behavior is elicited in a wide variety of different kinds of situations) and stability over time (e.g., similar behavioral tendencies are manifested in repeated encounters with the same kinds of trait-relevant situations). According to Miller, most of the participants in Baron's (1997) study come to the experiment with certain personality traits, including a cluster of causal psychological dispositions related to helping. Subjects are confronted with a relevant stimulus (e.g., seeing that a stranger has dropped a pen or being asked to make change for a dollar) and they perceive that a person is in need. The stranger's needs in the shopping mall experiment are not great, but the costs involved in helping are also low. On Miller's account, in some participants the situational stimulus triggers what Miller calls a "helping mechanism," which consists in a relevant set of mental dispositions that give rise to beliefs and desires (e.g., perhaps a belief that while the stranger's need is not great, the cost of helping is low, and a moderate desire to be helpful in such situations provided that it is not too inconvenient and that one is not in a bad mood) that can motivate helping behavior. The likelihood that a participant will help when given the opportunity is also partly a function of background conditions, such as a participant's mood, whether he or she is in a hurry, feeling guilty about something, and so on. Baron's study examines the impact of an environmental variable (e.g., ambient odor) on helping—an effect which appears to be mediated by the subject's mood. When contextualized within a larger psychological literature (e.g., Baron and Thomley 1994, who found that participants exposed to pleasant odors reported higher levels of positive affect than controls), the dominant explanation of what is going on in this shopping-mall study is that the fragrance lifts a person's mood, which in turn contributes to the activation of the person's helping mechanism and, in particular, a desire to help someone so as to maintain a good mood (Miller 2013, chap. 3).

[10] See, for example, Miller 2013, chapter seven, section three.

The evidence from psychological studies such as these at least arguably suggests that people's desires to help are often subject to certain kinds of morally problematic or ambiguous conditions (e.g., a desire to help when doing so will allow me to maintain my good mood). Such a disposition is not morally admirable in a way that qualifies it as a virtue, even to a minor degree. However, these conditions are not objectionable in a way that qualifies the character trait as a vice either. According to Miller, these traits are mixed from a moral perspective, in part because the frequency of expression of trait-relevant behavior depends significantly on the presence or absence of certain morally ambiguous, problematic, or irrelevant enhancers and inhibitors, and so do not quality as either traditional virtues or vices (see Miller 2013, chap. 1; 2014, chap. 2). So, the label "mixed" is appropriate.

Miller's Mixed Trait view can accommodate the oxytocin nasal-spray studies and apparently in just the same way that his view accounts for the fragrance cases—provided that we can identify some mental state dispositions triggered by the increase in oxytocin and that these psychological states play a genuine explanatory role in the causal explanation.[11] The presence of a situational variable, in this case the oxytocin nasal spray rather than the smell of freshly baked cookies, is correlated with, and indeed seems to causally influence, the frequency of certain kinds of behaviors used as experimental measures of trust or generosity. On Miller's account, in some subjects, the situational stimulus triggers a relevant set of mental dispositions that give rise to beliefs and desires that motivate the behavior. Among those participants for whom the character trait involved is such that, were the subject's oxytocin levels not so high, we would expect her generosity to decrease significantly

[11] It is crucial to Miller's account, and to many other virtue theoretic approaches to understanding human action in the moral domain that the behavior be explained at the psychological level in terms of character traits. At least for morally evaluable actions, the relevant mental state dispositions (e.g., beliefs and desires) must play a central causal role in manifesting the behavior. There would be ways of attempting to further specify the content of the mental states involved experimentally (see, for example, Kosfeld et al. 2005). Although Zak takes it that oxytocin release is "automatic and therefore beyond our conscious control," he thinks this triggers a mental state (a feeling of empathy, which perhaps involves both cognition—perceiving a situation from another's perspective, beliefs about the needs of others—and an emotional response to these needs, such as a desire to help) (Zak 2011, 223; Zak et al. 2007).

Because the Mixed Trait approach is cast in the terms of folk psychology, one might worry that it must assume a high degree of introspective access not just to our own mental states but that other people's motivations, intentions, beliefs, and desires are transparent enough for us to reliably discern them and use them to make predictions. If the approach did assume that, studies like Baron (1997) would pose a problem precisely because the influences of ambient odors on our willingness to help often, and perhaps typically, operate below the level of conscious awareness. Indeed, these operations are so subtle that it has required carefully controlled observations even to notice the influence of such factors. However, Miller allows that mental states can significantly influence our behavior below the level of conscious awareness (to see how Miller is thinking about the issue of introspective access see Miller 2013, chap., sec. 6).

(e.g., measurably, perhaps by about 17% in the same types of situations), the trait is mixed, not virtuous (other things being equal). So Miller's view can parse such cases nicely, apparently without the need for any ad hoc adjustments. Moreover, if the effects of oxytocin turn out to make the perspective of the other more psychologically salient, as Zak and colleagues (2007) suggest, Miller's view can accommodate these studies whether or not subjects are consciously aware of the dispositional mental states triggered by oxytocin. Without denying that oxytocin might also be important for moral behavior in other ways, in this respect its significance seems comparable to the influence of the smell of freshly baked cookies.

6. Conclusion

We are creatures of flesh and blood. So it should not shock us, nor would it have shocked ancients such as Aristotle or Hippocrates, to recognize that character traits like courage, benevolence, generosity, and self-control, or our moral evaluations themselves, are biologically mediated. We did not need contemporary neuroscience to tell us that traumatic brain injuries, like those of Phineas Gage, or the effects of aging can also result in long-term impairments in cognitive processes like memory, attention, and executive function or lead to significant changes in social behavior expressed by indifference or disinhibition (Milders et al. 2003). But the kind of information now becoming available and the level of detail at which we can study these processes will surely continue to shape moral theories in the times ahead in ways we cannot yet foresee. We have strong evidence that mood is an important influence on moral behavior, such as helping others (Miller 2013, chap. 3). We also know that mood is itself powerfully influenced by hormone and neurotransmitter levels, as are more stable features of temperament. Testosterone is associated with aggression and sexual desire. Pharmaceuticals like selective serotonin reuptake inhibitors (SSRIs) can affect mood (e.g., the SSRI Prozac is an antidepressant), characteristic emotional reactions, and other relatively stable features of one's psychology. The SSRI Paxil affects Big Five personality traits, lowering neuroticism and increasing extroversion (Tang et al. 2009), and lists of possible side effects include such cognitive effects as having suicidal thoughts. Other medications modulate tendencies toward anger, irritability, impulsivity, and anxiety.

Virtue theories or ordinary practices that appeal to character traits to explain and predict normatively evaluable behavior will eventually need to come to grips with the question of how such traits, or whatever mental dispositions might underlie them, relate to a detailed biological understanding of their causes. Though many empirical questions relevant to such an integration remain unanswered, recent neurobiological accounts of the associations between genes, molecular-level biochemical pathways,

complex social behaviors, emotions, and even moral judgment have led some to claim that such work has profound implications for morality and our ordinary understanding of persons. In this chapter, I have identified ways in which neurobiological accounts of the proximate causes of social behavior could, in principle, be relevant to virtue theory while resisting some of the hype that has surrounded recent research focusing on the neuropeptide oxytocin. I have also argued that, so far, the results of these studies are open to interpretations that are not nearly as revisionary for our commonsense understanding of persons or rational action as some would have us believe.

In the long years ahead, a broad scientific perspective on morality might well help us to understand the phylogenetic and ontogenetic origin and development of various capacities for social cognition and emotion that now get recruited in moral reaction and evaluation, how widely shared particular character or personality traits are, to what extent they are passed down culturally or via biology, the degree to which they can be shaped during an individual's moral development, what sort of role there is for learning or epigenetic influences, and how important certain key periods in life, such as childhood upbringing, are in shaping one's character. The project of responding to the situationist challenge has been a focus among many virtue theorists over the last decade. But another project awaits: thinking about how to integrate the neuroscience of moral psychology into virtue theory.[12]

References

Adams, R. M. (2006). *A Theory of Virtue.* Oxford: Oxford University Press.

Annas, J. (1993). *The Morality of Happiness.* Oxford: Oxford University Press.

Annas, J. (2006). "Virtue Ethics." In *The Oxford Handbook of Ethical Theory,* edited by D. Copp, 515–536. New York: Oxford University Press.

Annas, J. (2011). *Intelligent Virtue.* Oxford: Oxford University Press.

Anscombe, G. E. M. (1958). "Modern Moral Philosophy." *Philosophy: The Journal of the Royal Institute of Philosophy* 33: 1–19.

[12] This chapter was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. Thanks to John Doris, Thomas Hurka, Owen Flanagan, Adam Kadlac, Christian Miller, Adam Pelser, Walter Sinnott-Armstrong, Ray Yeo, and participants in the 2013 Character Project Summer Seminar for personal conversation and generous feedback on the topics related to this paper and to Christian Miller and Larry McKaughan for detailed comments on an earlier draft of the chapter.

Audi, R. (2001). "Epistemic Virtue and Justified Belief." In *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility,* edited by A. Fairweather and L. Zagzebski, 82–97. Oxford: Oxford University Press

Bales, K. L., Perkeybile, A. M., Conley, O. G., Lee, M. H., Guoynes, C. D., Downing, G. M., Yun, C. R., Solomon, M., Jacob, S., and Mendoza, S. P. (2013). "Chronic Intranasal Oxytocin Causes Long-Term Impairments in Partner Preference Formation in Male Prairie Voles." *Biological Psychiatry* 74 (3): 180–188.

Baron, R. (1997). "The Sweet Smell of . . . Helping: Effects of Pleasant Ambient Fragrance on Prosocial Behavior in Shopping Malls." *Personality and Social Psychology Bulletin* 23: 498–503.

Baron, R., and Thomley, J. (1994). "A Whiff of Reality: Positive Affect as a Potential Mediator of the Effects of Pleasant Fragrances on Task Performance and Helping." *Environment and Behavior* 26: 766–784.

Bartz J., Simeon D., Hamilton H., Kim S., Crystal S., Braun A., Vicens V., and Hollander E. (2011a). "Oxytocin Can Hinder Trust and Cooperation in Borderline Personality Disorder." *Social Cognitive and Affective Neuroscience* 6: 556–563.

Bartz J. A., Zaki J., Bolger N., and Ochsner K. N. (2011b). "Social Effects of Oxytocin in Humans: Context and Person Matter." *Trends in Cognitive Sciences* 15: 301–309.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). "Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans." *Neuron* 58: 639–650.

Bizman, A., Yinin, Y., Ronco, B., and Schachar, T. (1980). "Regaining Self-Esteem through Helping Behavior." *Journal of Psychology* 105: 203–209.

Churchland, P. M. (1981). "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78: 67–90.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain.* Cambridge, MA: MIT Press.

Churchland P. S. (2011a). *Braintrust: What Neuroscience Tells Us about Morality.* Princeton, NJ: Princeton University Press.

Churchland, P. S. (2011b). "How the Mind Makes Morals." Virginia Tech Carilion Research Institute. Distinguished Visiting Scholars Series. October 13. Video available at: http://research.vtc.vt.edu/videos/2011/oct/14/how-mind-makes-morals/.

Churchland, P. S., and Winkielman, P. (2012). "Modulating Social Behavior with Oxytocin: How Does It Work? What Does It Mean?" *Hormones and Behavior* 61: 392–399.

Crockett, M. J. (2012). "Beware Neuro-Bunk." TEDSalon London. November. Video available at: http://www.ted.com/talks/molly_crockett_beware_neuro_bunk.html.

Cushing, B. S., and Kramer, K. M. (2005). "Mechanisms Underlying Epigenetic Effects of Early Social Experience: The Role of Neuropeptides and Steroids." *Neuroscience and Biobehavioral Reviews* 29: 1089–1105.

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., Ten Velden, F. S., Van Dijk, E., and Feith, S. W. W. (2010). "The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict among Humans." *Science* 328: 1408–1411.

De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., and Handgraaf, M. J. J. (2011a). "Oxytocin Promotes Human Ethnocentrism." *Proceedings of the National Academy of Sciences* 108: 1262–1266.

De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., and Handgraaf, M. J. J. (2011b). "Reply to Chen et al.: Perhaps Goodwill Is Unlimited, but Oxytocin-Induced Goodwill Is Not." *Proceedings of the National Academy of Sciences* 108: E46.

DeYoung, C. G. (2010). "Personality Neuroscience and the Biology of Traits." *Social and Personality Psychology Compass* 4 (12): 1165–1180.

Doris, J. (1998). "Persons, Situations, and Virtue Ethics." *Noûs* 32: 504–530.

Doris, J. (2002). *Lack of Character: Personality and Moral Behavior.* Cambridge: Cambridge University Press.

Doris, J. (2010). "Heated Agreement: Lack of Character as Being for the Good." *Philosophical Studies* 148: 135–146.

Flanagan, O. (2009). "Moral Science? Still Metaphysical after All These Years." In *Moral Personality, Identity and Character: Explorations in Moral Psychology.* Edited by D. Narvaez and D. K. Lapsley, 52–78. Cambridge: Cambridge University Press.

Fleeson, W. (2001). "Toward a Structure- and Process-Integrated View of Personality: Traits as Density Distributions of States." *Journal of Personality and Social Psychology* 80: 1011–1027.

Fleeson, W., and Gallagher, P. (2009). "The Implications of Big Five Standing for the Distribution of Trait Manifestation in Behavior: Fifteen Experience-Sampling Studies and a Meta-Analysis." *Journal of Personality and Social Psychology* 98: 1097–1114.

Fleeson, W., and Noftle, E. (2008). "The End of the Person-Situation Debate: An Emerging Synthesis in the Answer to the Consistency Question." *Social and Personality Psychology Compass* 2: 1667–1684.

Francis, D., Diorio, J., Liu, D., and Meaney, M. J. (1999). "Nongenomic Transmission across Generations of Maternal Behavior and Stress Responses in the Rat." *Science* 286: 1155–1158.

Francis, D. D., Young, L. J., Meaney, M. J., and Insel, T. R. (2002). Naturally Occurring Differences in Maternal Care Are Associated with the Expression of Oxytocin and Vasopressin (V1a) Receptors: Gender Differences. *Journal of Neuroendocrinology* 14: 349–353.

Frankfurt, H. G. (2004). *The Reasons of Love.* Princeton, NJ: Princeton University Press.

Fries A. B., Ziegler T. E., Kurian J. R., Jacoris S., and Pollak S. D. (2005). "Early Experience in Humans Is Associated with Changes in Neuropeptides Critical for Regulating Social Behavior." *Proceedings of the National Academy of Sciences* 102 (47): 17237–17240.

Gray, K., Young, L., and Waytz, A. (2012a). "Mind Perception Is the Essence of Morality." *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory* 23 (2): 101–124.

Gray, K., Waytz, A., and Young, L. (2012b). "The Moral Dyad: A Fundamental Template Unifying Moral Judgment." *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory* 23 (2), 206–215.

Griffiths, P. (1997). *What Emotions Really Are.* Chicago: University of Chicago Press.

Griffiths, P. (2004). "Emotions as Natural and Normative Kinds." *Philosophy of Science* 71: 901–911.

Guéguen, N. (2012). "The Sweet Smell of . . . Implicit Helping: Effects of Pleasant Ambient Fragrance on Spontaneous Help in Shopping Malls." *Journal of Social Psychology* 152 (4): 397–400.

Harman, G. (1999). "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99: 315–331.

Harman, G. (2000). "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society* 100: 223–226.

Harman, G. (2001). "Virtue Ethics without Character Traits." In *Fact and Value,* edited in A. Byrne, R. Stalnaker, and R. Wedgewood, 117–127. Cambridge, MA: MIT Press.

Harman, G. (2003). "No Character or Personality." *Business Ethics Quarterly* 13: 87–94.

Harman, G. (2009). "Skepticism about Character Traits." *Journal of Ethics* 13: 235–242.

Harris, M., and Smith, R. (1975). "Mood and Helping." *Journal of Social Psychology* 91: 215–221.

Hursthouse, R. (1999). *On Virtue Ethics.* Oxford: Oxford University Press.

Job, R. (1987). "The Effect of Mood on Helping Behavior." *Journal of Social Psychology* 127: 323–328.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). "Oxytocin Increases Trust in Humans." *Nature* 435: 673–676.

Levin, P., & Isen, A. (1975). "Further Studies on the Effect of Feeling Good on Helping." *Sociometry* 38, 141–147.

MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory.* 3rd ed. Notre Dame, IN: University of Notre Dame Press.

McGuire, B. (1988). "Effects of Cross-Fostering on Parental Behavior of Meadow Voles (*Microtus pennsylvanicus*)." *Journal of Mammalogy* 69 (2), 332–341.

McKaughan, D. J. (2012). "Voles, Vasopressin, and Infidelity: A Molecular Basis for Monogamy, a Platform for Ethics, and More?" *Biology and Philosophy* 27: 521–543.

McKaughan, D. J., and Elliott, K. C. (2012). "Voles, Vasopressin, and the Ethics of Framing." *Science* 338 (6112): 1285.

McKaughan, D. J., and Elliott, K. C. (2013). "Backtracking and the Ethics of Framing: Lessons from Voles and Vasopressin." *Accountability in Research* 20 (3): 206–226.

Melchers, M., Montag, C., Markett, S., and Reuter, M. (2013). "Relationship between Oxytocin Receptor Genotype and Recognition of Facial Emotion." *Behavioral Neuroscience* 127 (5): 780–787.

Milders, M., Fuchs, S., and Crawford, J. R. (2003). "Neuropsychological Impairments and Changes in Emotional and Social Behaviour Following Sever Traumatic Brain Injury." *Journal of Clinical & Experimental Neuropsychology,* 25(2), 157–172.

Miller, C. (2013). *Moral Character: An Empirical Theory.* Oxford: Oxford University Press.

Miller, C. (2014). *Character and Moral Psychology.* Oxford: Oxford University Press.

Mischel, W. and Shoda, Y. (1995). "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure." *Psychological Review* 102: 246–268.

Montag C, Fiebach C. J., Kirsch P., and Reuter M. (2011). "Interaction of 5-HTTLPR and a Variation on the Oxytocin Receptor Gene Influences Negative Emotionality." *Biological Psychiatry* 69: 601–603.

Moran, J. M., Young L. L., Saxe R., Lee S. M., O'Young D., Marvos P. L., and Gabrieli, J. D. (2011). "Impaired Theory of Mind for Moral Judgment in High-Functioning Autism." *Proceedings of the National Academy of Sciences* 108 (7): 2688–2692.

Rorty, R. (1965). "Mind-Body Identity, Privacy, and Categories." *Review of Metaphysics* 19: 24–54.

Russell, D. (2009). *Practical Intelligence and the Virtues.* Oxford: Clarendon Press.

Slote, M. (1992). *From Morality to Virtue.* Oxford: Oxford University Press.

Snow, N. (2010). *Virtue as Social Intelligence: An Empirically Grounded Theory.* New York: Routledge.

Shamay-Tsoory, S. G., Fischer, M., Dvash, J., Harari, H., Perach-Bloom, N., and Levkovitz, Y. (2009). "Intranasal Administration of Oxytocin Increases Envy and Schadenfreude (Gloating)." *Biological Psychiatry* 66 (9): 864–870.

Stich, S. (1983). *From Folk Psychology to Cognitive Science.* Cambridge, MA: MIT Press.

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., Shelton, R., and Schalet, B. (2009). "Personality Change during Depression Treatment: A Placebo-Controlled Trial." *Archives of General Psychiatry* 66 (12):1322–1330.

Urmson, J. O. (1988). *Aristotle's Ethics.* Oxford: Blackwell.

Walter, N., Montag, C., Markett, S., Felten, A., Voigt, G., and Reuter, M. (2012). "Ignorance Is No Excuse: Moral Judgments Are Influenced by a Genetic Variation on the Oxytocin Receptor Gene." *Brain and Cognition* 78: 268–273.

Wang, H., Duclot, F., Liu, Y., Wang, Z. and Kabbaj, M. (2013). "Histone deacetylase Inhibitors Facilitate Partner Preference Formation in Female Prairie Voles." *Nature Neuroscience* 16 (7): 919–924.

Weyant, J. (1978). "Effects of Mood States, Costs, and Benefits of Helping." *Journal of Personality and Social Psychology* 36: 1169–1176.

Winter, M. (2011). *Rethinking Virtue Ethics.* Dordrecht: Springer.

Young, L. J., and Flanagan-Cato, L. M. (2012). "Editorial Comment: Oxytocin, Vasopressin and Social Behavior." *Hormones and Behavior* 61: 227–229.

Young, L., and Saxe, R. (2008). "The Neural Basis of Belief Encoding and Integration in Moral Judgment." *NeuroImage* 40: 1912–1920.

Young, L., and Saxe, R. (2009). "Innocent Intentions: A Correlation between Forgiveness for Accidental Harm and Neural Activity." *Neuropsychologia* 47: 2065–2072.

Young, L., Camprodon, J., Hauser, M., Pascual-Leone, A., and Saxe, R. (2010). "Disruption of the Right Temporoparietal Junction with Transcranial Magnetic Stimulation Reduces the Role of Beliefs in Moral Judgments." *Proceedings of the National Academy of Sciences* 107: 6753–6758.

Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). "The Neural Basis of the Interaction between Theory of Mind and Moral Judgment." *Proceedings of the National Academy of Sciences* 104 (20): 8235–8240.

Young, L., Koenigs, M., Kruepke, M., & Newman, J. P. (2012). "Psychopathy Increases Perceived Moral Permissibility of Accidents." *Journal of Abnormal Psychology* 121: 659–667.

Zak, P. J. (2005). "Trust: A Temporary Human Attachment Facilitated by Oxytocin." *Behavioral and Brain Sciences* 28: 368–369.

Zak, P. J. (2011). "Moral Markets." *Journal of Economic Behavior & Organization* 77 (2): 212–233.

Zak P. J. (2012a). *The Moral Molecule: The Source of Love and Prosperity.* New York: Dutton.

Zak, P. J. (2012b). Interview with Gary Rosen. "Can I Trust You? A Molecule May Decide," *Wall Street Journal*, April 27. Video available at: http://online.wsj.com/article/SB10001424052702304811304577365782995320366.html.

Zak, P. J., Kurzban, R., and Matzner, W. T. (2005). "Oxytocin Is Associated with Human Trustworthiness." *Hormones and Behavior* 48: 522–527.

Zak, P. J., Stanton, A. A., Ahmadi, S. (2007). "Oxytocin Increases Generosity in Humans." *PLoS ONE* 2 (11): e1128.

Zak, P. J., Park, J., Ween, J. Graham, S. (2006). "An fMRI Study of Interpersonal Trust with Exogenous Oxytocin Infusion." *Society for Neuroscience*, Program no. 2006-A-130719-SfN.

29 }

Character and Coherence: Testing the Stability of Naturalistically Observed Daily Moral Behavior

Matthias R. Mehl, Kathryn L. Bollich, John M. Doris, and Simine Vazire

Consistency, Stability, and Coherence

The virtue ethics and situationism debate that has lately perturbed philosophical ethics is, in large measure, a transposition of the person-situation debate that divided personality and social psychology during the 1960s, 1970s, and 1980s (Donnellan, Lucas, & Fleeson, 2009; Epstein, 1979; Kenrick & Funder, 1988; Miller, 2014; Mischel, 1968). "Character skeptics" in philosophy argue that behavioral inconsistency problematizes notions of robust character traits, such as virtues, much as "situationist" personality and social psychologists argue that behavioral inconsistency undermine traditional notions of personality traits.

In psychology, peace has fallen on the front, with even diehard personologists (e.g., Funder, 2009) acknowledging that situations matter, and even unrepentant situationists (e.g., Mischel, 2009) giving personality variables their due. In short, "everybody's an interactionist," (i.e., acknowledges that both the person and situation matter and that they depend on each other) and, perhaps, they have always been so (Doris, 2002: 25–26). While the occasional reactionary ruins a colloquium dinner or question-and-answer period by reliving the bad old days, the spirit of personality and social psychology today strives to be ecumenical, as many research programs deploy *both* personal and situational variables (e.g., Cameron, Payne, & Doris, 2013; Fleeson, 2012; Fleeson & Noftle, 2008; Leikas, Lönnqvist, & Verkasalo, 2012; Sherman, Nave, & Funder, 2010), and systematic differences between (primarily) social psychologists and (primarily) personality psychologists exist

more in their research methods than their research content (Tracy, Robins, & Sherman, 2009).

In this chapter, we celebrate today's irenic spirit, reporting research we hope will clarify thinking on the person-situation interface for both psychologists and philosophers. But before doing so, we'll need to trace some outlines of the classic debate. We'll follow convention in saying that *stability* is the tendency to repeatedly produce similar behaviors in iterated trials of similar situations: if you give to the holiday charity drive at your office every year without fail, you exhibit a stable charitable behavior. We'll likewise follow convention in treating *consistency* as the tendency to repeatedly produce similar behaviors across a range of differing situations: if you give to the office charity drive, but also to charitable solicitations you receive in the mail and in response to the entreaties of panhandlers who approach you on the street, your charitable behavior exhibits consistency. The distinction, of course, is permeable: time changes all, so iterated situations are never exactly similar (meaning that stability may imply a measure of consistency), and human lives exhibit considerable continuity (meaning that consistency may imply a measure of stability). Nonetheless, situations vary tremendously with respect to their similarity and dissimilarity, and the stability/consistency distinction helps to make theoretical sense of this fact (see Furr, 2009b; Ozer, 1986; for a comprehensive discussion of 36 different consistency concepts filling a matrix created by three determinants of behavior, three definitions of similarity, and four definitions of behavioral enactment, see Fleeson and Noftle, 2008).

The difficulty that motivated situationists and character skeptics was the appearance of surprising behavioral *in*consistency: people often behave dissimilarly even across situations that might be expected to engage the same disposition: you might give at the office but not through the mail or help out the panhandler but not a sibling who asks for a loan. But if this is the problem, it's not yet obvious what the problem is. Consider charity (we've chosen our example with malice aforethought): even a very charitable person doesn't give to charity on every charity-relevant occasion, nor should she be expected to do so. If you did, you'd rapidly become a destitute person, and not able to engage in behaviors characteristic of the charitable person. Charitable behavior, quite appropriately, is constrained by a person's means (the practice of tithing recognizes this, by suggesting a percentage, rather than an amount). But there's something else: however much money you have, your patterns of charity will also, quite appropriately, be structured by conviction and interest: if you think college is a scam, you won't be giving to higher education.

Although not all inconsistency is so readily explicable (a fact skeptics and situationists will be anxious to emphasize), we've now noted a critical point: *behavior can be orderly and intelligible without being highly consistent.* Your charitable habits, though quite inconsistent (say, you're exactly as likely to say no as yes to the full range of solicitations), make perfect sense given your

circumstances and values. We'll call this kind of making sense *coherence*: a person is possessed of a coherent personality to the extent that her overall pattern of behavior may be seen as orderly and intelligible, with respect to the sort of person she is hypothesized to be (thus, we construe coherent *personality* as effecting some measure of consistency and stability in *behavior*; conversely, we can treat the latter as *evidence* for the former.). This observation can be understood in terms of a distinction between "nominal" versus "psychological" situations: nominal situations are circumstances that can be characterized "objectively," more or less independently of the psychological particulars of those occupying them—e.g., an academic lecture—whereas psychological features are characterized "subjectively," with reference to the psychological saliences and effects of the situation on the person—e.g., being bored to the point of hot, bitter tears (Doris, 2002: 76–77; Shoda, Mischel, & Wright, 1994).

It is tempting to think that this observation makes a sufficient response to situationism and character skepticism. For surely, focusing on the nominal at the expense of the psychological may cause us to miss much coherence in personality, with the result that the skeptical difficulty becomes exaggerated. When we attend to the fact that Jones found one lecture boring and another not, the fact that he dozed through one and listened raptly to the other is hardly a confounding inconsistency. Yet the person-situation debate was not, contra the declarations of some observers (Annas, 2005: 637), much ado about nothing. If you care about coherence, you should care about nominal stability and nominal consistency, for both contribute to coherence.

Most conspicuously, departures from nominal stability and consistency may result in failures of coherence. As inconsistency increases, intelligibility and order likely decrease: at the theoretical limit, if a person's behavior exhibited no consistency at all across situations, it wouldn't make much sense. As for consistency, so for stability: a complete lack of association between a person's behavior at one time and the next is not the recipe for order and intelligibility. This is so even if we posit total psychological stability and consistency: if a person's behavior wildly fluctuates—is "at chance"—for objectively observable behaviors that are criterial for a trait, he is less than maximally coherent with regard to that trait (for discussion, see Doris, 2002: 76–85). If Jones is as likely, when he passes the Salvation Army bucket during the holidays, to toss in some change, kick it over, organize a game of soccer using the bucket as a ball, or a myriad of other things we are not clever enough to envisage, he is not maximally coherent with regard to charitable behavior, even if he is psychologically consistent—say he's dedicated his life "to keeping those simpering do-gooders on their toes."

Happily, most people aren't put together like Jones, and most people aren't completely patternless. Some degree of coherence in personality, surely, is overwhelmingly the rule. But how to confirm the postulated rule? This raises the question of how to "measure" personality. The question is not an innocuous

one, since in personality, as elsewhere, to measure is not only to observe, but also to manufacture. What you seek influences what you will find.

To make this concrete, look first to the virtue ethics and situationism debate in philosophy. Character skeptics focused on "overt behavior" (in the sense of behavior that is "out there" in the world); the pervasiveness of behavioral inconsistency was supposed to motivate skepticism about cross-situationally consistent ("robust") traits, as exemplified by the virtues. Some defenders of virtue ethics complained that skeptics focused "overwhelmingly on behavioral dispositions" while "largely ignor[ing]" internal psychological factors that are widely believed to be crucial elements of moral functioning (Adams, 2006: 121, 136; cf. Annas, 2005: 638; Swanton, 2003: 30–31).[1] According to Adams (2006: 132), dispositions involved in virtue "will typically be dispositions, first of all, to types of psychological processing, rather than directly to types of observable physical behavior."

In effect, the philosophers were suggesting giving greater attention to psychological situations. As we've already intimated, this strategy is not without difficulty. There's risk of unfalsifiability, since de-emphasizing overt behavior in favor of internal states might insulate attributions of virtue from observational testing (Doris, 2002: 26; 2005: 664). There's also ethical difficulty: if behavior didn't matter, good intentions would count as much as good deeds, and we'd have to find another way to pave the road to hell. Defenders of virtue, then, should concede that dispositions to virtuous behavior are necessary for virtue, even if they are not sufficient (Adams, 2006: 121).

If both inner and outer matter, how are they supposed to relate? A natural thought is that internal coherence is reflected in a kind of external order. The virtuous person, such as Aristotle's exemplary *phronimos*, may sometimes tell it straight and sometimes shade the truth, or sometimes stand his ground and sometimes run like a thief; but this behavioral variety is structured by orderly patterns of cognition and motivation—say, by an abiding conception of what makes a life good. Perhaps the *phronimos* makes his end a life rich in emotionally intimate relationships: sometimes such a life is facilitated by being bluntly honest, and other times not. And so for other goods: a life of martial valor sometimes requires the ultimate sacrifice and, at other times, requires fighting another day. But this nominal inconsistency does not entail incoherence; reference to the actor's understanding of martial valor, for instance, renders her martial conduct explicable.

The problem, unfortunately, is that the inner and the outer aren't always, or often, so neatly related, as we can see by returning to personality psychology. Like virtue ethicists, some personality psychologists de-emphasized overt

[1] In fact, the importance of inner states was quite explicitly acknowledged by skeptics (Doris 2002: 17, cf. 86–89; 2005: 663).

behavior (Shoda et al., 1994). Complaints about unfalsifiability are again relevant here (Mischel, 1968: 56; Kenrick and Funder, 1988: 24), but a more fertile question is: supposing behavior is not a definitive window on personality, what is?

Historically, the answer has been "paper and pencil" personality assessment based on *self-reports*, where investigators have found regularities in test responses favorable to standard theoretical constructions of personality (Holzman and Kagan, 1995: 5–8; Ross and Nisbett, 1991: 98). Without doubt, self-reports are of crucial interest; at a minimum, they tell us quite a lot about how people view themselves, and certainly, how people view themselves has quite a lot to do with what they are like.

The bad news is that how people view themselves may diverge pretty sharply from other perspectives on their personality. Although self-views are not completely untethered to reality (Vazire & Carlson, 2010, 2011), they have their shortcomings, due to biases or blindspots in evaluations of oneself (Vazire, 2010). In other words, people are often motivated to see themselves in a particular way and, at times, are missing information that would allow them to have an accurate impression of themselves. In general, people are motivated to be positively biased (Alicke & Sedikides, 2009; Sedikides & Alicke, 2012). That is, people are more likely to rate themselves as better—for example, more charitable and fair, and as possessing better driving and teaching abilities—than their peers (Alicke, 1985; Cross, 1977; Epley & Dunning, 2000; Messick, Bloom, Boldizar, & Samuelson, 1985; Svenson, 1981). Even prisoners rate themselves as more prosocial than the average prisoner and member of the community (Sedikides, Meek, Alicke, & Taylor, 2013). On the other hand, some people are biased in the opposite direction, and are more prone to overly negative evaluations of themselves (Kwan, John, Kenny, Bond, & Robins, 2004). Biases aside, people are also not always in the position to gain the necessary information to make an accurate assessment of themselves. For example, our line of vision does not allow us to see the way we carry ourselves or our natural facial expressions (Albright & Malloy, 1999). Taken together, we have reason to be wary of putting our total trust in the accuracy of self-reports.

With that said, the accuracy of one's self-ratings isn't always off. Indeed, self-knowledge is quite dependent on the specific trait or behavior of interest. People are generally very good at judging their more internal traits, such as those related to anxiety and neuroticism (Vazire, 2010). So, if we're interested in how nervous someone is before they give a presentation, we can probably trust their personal report. However, people are not very accurate at judging traits that are more evaluative, such as intelligence and creativity, so if we want to know if our presenter wowed with her smarts, we better ask the audience (John & Robins 1993; Vazire, 2010). In fact, close others are actually more accurate in rating observable, evaluative traits. It's not a stretch, then, to

assume that if people are not very accurate in judging their evaluative traits like intelligence, then they are also not very accurate in rating their moral behaviors.

Of special concern here is the relationship between self-report and overt behavior, or in philosophical terms, the relation between the internal and the external. Unfortunately, the relationship between self-report measures and overt behavior is often not very strong (Gosling, John, Craik, Robins, 1998; Holzman and Kagan, 1995: 5; Pervin, 1996: 315; McClelland, 1985: 818–820; Mischel, 1968: 25; Ross and Nisbett, 1991: 98–99). In short, how people see themselves—or at least, what they report about themselves—often proves a limited guide to what they actually do. For example, when examining objective, naturalistic measures of actual—relatively mundane—behaviors and their relation to self-reports of those same behaviors, researchers found they correlated, on average, .23 (Vazire & Mehl, 2008). Whether .23 is a small or large correlation depends on your expectations, but it certainly cannot be considered abundant evidence of self-knowledge. Gosling and his colleagues found similar, though maybe slightly larger, self-report-behavior correlations in a lab context (Gosling et al., 1998).

Vazire's (2010) self-other knowledge asymmetry (SOKA) model provides a framework for understanding and predicting the bright spots and blind spots in people's self-knowledge. According to the SOKA model, self-views are likely to be most accurate for internal, evaluatively neutral traits, and least likely to be accurate for behavioral, evaluative traits. That is, the degree to which people exhibit self-insight depends on whether you are asking about their thoughts and feelings or their behaviors, and whether the trait you are asking about is neutral or evaluative (i.e., very desirable or undesirable).

Regarding behavioral consistency, the situation looks to be this: self-reports register substantial consistency, but overt behavioral measures less so. Accordingly, it has sometimes looked as though there are two discrete disciplines, one of self-report taxonomies and one of overt behavioral measures, with very different standards of success and, accordingly, very different results (see Epstein, 1983: 361). And if the gold standard of personality is consistency, you might think that personality psychologists are well served by sticking with self-reports.

We don't think so. Like many others, we think the best approach to personality is multipronged, involving self-report, peer report, behavioral observation, and more (Baumeister, Vohs & Funder, 2007; Furr, 2009a; Vazire & Mehl, 2008). Nor do we believe that personality psychologists should fear overt behavior. The reason personality theorists have no reason to fear behavior is that systematic behavioral observation, of the sort we are about to describe, gives evidence of coherence. This coherence is not uncovered by investigation of psychological situations but by the observation of behavior in what are best understood as nominal situations, coded by third-party observers. Our

strategy in the face of character skepticism is not to retreat, as others have done, to the hidden citadel of psychological situations, but to meet situationists on their home field, the observation of behavior in nominal situations.

Behavior and Measurement

In many of psychology's neighboring disciplines, including anthropology, sociology, and primatology, the observation of subjects in their natural habitat is a common practice, held in high methodological regard. Yet naturalistic observation has a surprisingly thin history in psychology, despite some notable exceptions. For example, Roger Barker and Herbert Wright spearheaded a series of ecological observational studies on "behavior settings," including the famous case study *One Boy's Day* (1951). Inspired by Barker and Wright, Kenneth Craik (2000) developed what he called the *lived day analysis*, a series of case studies in which a research team followed participants for an entire day with a video camera. Researchers at the Center on Everyday Lives of Families, at the University of California, Los Angles, embarked on a large, interdisciplinary, naturalistic observation project to study family dynamics in the real world (Campos, Graesch, Ochs, Repetti & Bradbury, 2009). This is hardly an exhaustive review; some naturalistic observation work has also been conducted in developmental psychology (e.g., Tudge, Hogan, & Etz, 1999) and social psychology (e.g., Fraley & Shaver, 1998). Yet there remains a substantial gap between the public idea—and scientific ideal—of psychologists as professional observers of human behavior and the way much psychological science actually gets done.

Obviously, in contrast to primatologists, psychologists are in the privileged position to be able to obtain good data simply by asking their subjects—an inarticulate person would make an eloquent bonobo (Rozin, 2001). In addition to this pleasing directness, self-reports are far more economical than obtaining behavioral observation data. But, as we've said, the limitations of self-report assessment are considerable. If the goal is to develop a comprehensive understanding of human behavior, we must complement questionnaire-based assessments with observational methods. Just as experience sampling methods[2] were developed to bring self-report data collection to where experiences naturally happen (Conner, Tennen, Fleeson, & Barrett, 2009), naturalistic observation methods are needed to bring behavioral data collection to where behavior naturally occurs.

[2] In experience sampling studies, participants are prompted several times per day to complete a short questionnaire on their momentary thoughts, feelings, and behavior. Experience sampling methods were developed to collect "first-hand," in vivo, real-time information on people's daily thoughts, feelings, and behavior (Conner et al., 2009).

To this end, Mehl and colleagues have developed the Electronically Activated Recorder (EAR) as an "observational ecological momentary assessment tool" (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001; Mehl, Robbins, & Deters, 2012). The basic idea behind the EAR is simple: to obtain a sense of who people are and what they do by listening in on sound bites of their moment-to-moment social interactions. Technologically, the EAR is an audio recorder that periodically samples brief snippets of ambient sounds while participants go about their lives. Conceptually, it is an unobtrusive observation method that yields acoustic logs of people's lived days as they naturally unfold.

In early versions, the EAR consisted of a microcassette recorder and a microchip that triggered the recordings, with the chip glued to the recorder and soldered wires connecting the two. It was quite a delicate system—and participants had to remember to flip the tape after about 24 hours so that two days worth of recordings would fit onto a 45-minute microcassette. Fortunately, technology has advanced. Currently in use is a fourth-generation EAR system, the iEAR, that consists of a free, Apple-certified iOS app that runs on iPod touch (and iPhone) devices. To minimize privacy concerns, the recordings are stored locally on the device and are transferred to the researchers' computer at the end of the study, after participants have had an opportunity to review them and arrange for the redaction of any material they deem sensitive.[3] With the iEAR, we now have a reliable and affordable system that allows for the naturalistic (acoustic) observation of people's daily social environments and interactions.

Given the novelty of the method, the research focused initially on the psychometrics of EAR-assessed daily social behavior. In a first series of studies, the method was used to show (a) that a broad spectrum of behaviors can be assessed reliably and with low levels of reactivity from the sampled ambient sounds (Mehl & Holleran, 2007), (b) that these behaviors show large between-person variability and good temporal stability (Mehl & Pennebaker, 2003), and (c) that they show a moderate degree of empirical correspondence to theoretically-related measures such as the Big Five personality traits (Mehl et al., 2006), subclinical depression (Mehl, 2006), and narcissism (Holtzman, Vazire, & Mehl, 2010).

As the value of a novel method ultimately hinges on its ability to address questions that cannot be addressed otherwise, second generation studies have focused on the EAR's potential to (1) provide real-world measures of constructs that are independent of self-report and (2) allow for the assessment of subtle social behaviors that evade self-report (Mehl & Robbins, 2012). For example,

[3] Participants appreciate the opportunity to review their recordings; only rarely, though, do they actually delete files. In our studies, participants tend to delete less than 0.01% of the sound files, suggesting that the provided files provide a good approximation of their actual activities and interactions (Mehl & Holleran, 2007).

in a binational EAR study (Ramirez-Esparza, Mehl, Álvarez-Bermúdez, & Pennebaker, 2009), it was found that although American participants reported being *more* sociable than their Mexican counterparts, they spent *less* time with others and had *fewer* social (i.e., non-instrumental) conversations. Intriguingly, whereas Americans rated themselves significantly *higher* than Mexicans on the item "I see myself as a person who is talkative," they spent almost 10% *less* time talking.

In a similar way, the EAR was used to debunk the long-standing myth that women are more talkative than men (Mehl, Vazire, Ramírez-Esparza, Slatcher, & Pennebaker, 2007). A re-analysis of six data sets showed that both sexes used on average about 16,000 words per day. The estimated gender difference of 546 words compared to a range of over 46,000 words between the least talkative and most talkative individual (695 vs. 47,016) rendered significance testing close to meaningless—and spoke powerfully to the magnitude of individual differences.

Yet another study tested the extent to which subtle, affect-associated social behaviors follow diurnal rhythms (Hasler, Mehl, Bootzin & Vazire, 2008). Consistent with the idea that positive affect but not negative affect is subject to circadian regulation, the behavioral expression of positive (laughing, singing) but not negative (arguing, sighing) affect showed periodic trends. Together, these studies establish the validity of the EAR as an unobtrusive observation methodology and speak to its "added value" for assessing subtle yet psychologically important social behaviors that tend to evade self-report. In the present chapter, we report on the application of EAR methodology to moral behavior, and articulate some lessons for the study of character.

Behavior, Stability, and Morality

In the research reported here (for a more detailed description of the employed methodology and obtained results, please see Bollich, Doris, Vazire, Raison, Jackson, & Mehl, under review), we used the EAR to directly test the stability of individual differences in actual everyday moral behavior. Using this method allowed us to (a) objectively assess actual behaviors and (b) sample from a broad range of a person's daily life, maximizing the generalizability of research.

The first step in our research was determining what behaviors we could code from the EAR that constitute everyday moral (or immoral) behavior. Although philosophical attempts to define morality have met with limited success (Nado, Kelly, & Stich, 2009), we simply operationalized moral behavior as behavior with a prosocial (or antisocial) component. While one can certainly argue about which behaviors are moral and how moral they are (more on this below), it seems that many uncontroversial examples of moral conduct (e.g., giving to charity) involve prosocial behavior, while many uncontroversial

examples of immoral behavior (e.g., stealing from children) involve antisocial behavior. Whatever else "the domain of morality" involves, it very frequently involves the social, so investigating prosocial behavior is a theoretically well-motivated starting point for empirical work on the coherence of moral behavior. This approach, of course, meant that we were limited to observations assessing nominal situations; but as we have said, behavior in nominal situations is a critical component of personological coherence, and one central in the person-situation debate, so this limitation may actually be seen as a strength of the current research.

To make our start, we sought to identify behaviors that are (1) acoustically detectable and (2) occur with some regularity in everyday life. This excluded rare behaviors, such as rescuing a drowning child, and emphasized interpersonal behaviors, which are more likely to be acoustically detectable. We also sought to cover a range of everyday behaviors whose presence, like showing sympathy, is conventionally associated with morality. The behaviors we examined are presented in table 29.1, as well as examples of each behavior. Obviously this methodology, as with many methodologies, does not allow us to address the psychological antecedents of moral behavior, so we are limited to behaviors that are morally relevant overt behaviors; again, we are limited to nominal situations. We are therefore not able to give a complete picture of moral functioning, but we are able to develop a necessary part of the picture.

TABLE 29.1 } **Examples of Moral Daily Behaviors**

Moral Daily Behaviors	**Examples (adapted from actual EAR codings)**
Showing gratitude	*"Thank you so much! I really appreciate it. You helped me a lot!"*
Showing affection	*"I love you; I really do."*
Offering praise, making compliments	*"You're learning a lot. You are doing great!"*
Offering help, support	*"Can I carry that for you?"*
Apologizing	*"I am so sorry; I didn't mean to make you look bad."*
Expressing hope, showing optimism	*"I really want to be proactive about that. It'll be great."*
Showing sympathy, empathy, concern	*"I don't know why she did such hurtful things to you."*
Blaming	*"I came to see you and you were gone! You made me wait an hour!"*
Being sarcastic	*"You want my lemonade?? Sure! After you drank the whole thing!"*
Being condescending or arrogant	*"How did you get so retarded? Oh, yes. I know what it means."*
Complaining or whining	*"You make it so difficult!"*
Criticizing	*"Then get off your butt and fix them lunch."*
Bragging	*"My dad's car collection is so large he had to build a new garage."*
Expressing pessimism	*"I doubt this medication will help"*

For this project, all sound files were coded for all behaviors listed in table 29.1 (more details on the coding process are provided later in the chapter). We measured stability of moral behaviors in two ways. First, we assessed *momentary stability*. To do this we separated odd files from even files (for the time period of the monitored weekend), aggregated the two sets separately for each participant, and then correlated those two sets. This measure of stability is similar to Epstein's (1979, 1980) early work on trait and behavior stability that sought to measure stability in a variety of ways. By correlating averaged behaviors from odd files and even files, this gave us a measure of momentary stability throughout a few days. Second, following traditional longitudinal work, we also assessed *rank-order* (or *test-retest) stability* by correlating averaged behaviors from one weekend with those from a second weekend. This correlation indicates the likelihood that someone who is particularly compassionate, helpful, and affectionate on one weekend, will be particularly compassionate, helpful, and affectionate on a following weekend.

An obvious difficulty in testing whether individual differences in moral behavior are stable is determining a benchmark for stability. How much temporal stability is necessary to say that behavior is stable? Given that the interpretation of effect size can be somewhat subjective, we opted to compare the temporal stability of moral behaviors with another, relatively stable construct: language use. Linguistic variables provide a particularly strong test for several reasons. For one, language can be coded very easily and reliably from EAR files using computerized text analysis (Mehl, Gosling, & Pennebaker, 2006), which means that stability estimates will not suffer from a lack of reliability. Many linguistic variables are also quite neutral (e.g., article and preposition usage), ensuring that their occurrence in one's daily life should be unrelated to our moral constructs of interest, providing an independent measure of the stability in a person's life. Finally, linguistic variables can be matched to the coded moral behaviors with respect to their base rates. For example, if the moral behavior of gratitude happens, on average, in 3% of people's conversations, its stability can be compared against a linguistic behavior that occurs with approximately equal frequency (e.g., use of articles). This is important because the base rate and reliability of a behavior are inextricably linked (Epstein, 1979).

A priori considerations also favor use of linguistic stability as a benchmark. For example, a carpenter might be expected to repeatedly use words like *saw* in her work, just as a doctor might be expected to repeatedly use words like *patient*: intuitively, we'd expect the same words to appear fairly stable in the course of a person's activities (for consistency, perhaps not so much: the carpenter might seldom say *hammerdrill* in church, but frequently on the job site, the doctor might seldom say *fever* on dates, but frequently on her rounds, and so on.). Although these content words (e.g., fever, hammerdrill) may be more prone to the influence of situations, function words, or words that are content-free (e.g., articles, prepositions, and numbers) should

be an especially consistent construct. Indeed, past research indicates that function words can be excellent markers of individual differences (Pennebaker & King, 1999) and provides evidence of substantial cross-situational consistency (Mehl & Pennebaker, 2003). Although cross-situational consistency is not the same as cross-temporal stability, as we discussed earlier, the two are related. A construct with high cross-situational consistency (e.g., function words) is also likely to have high levels of stability. Thus, by comparing the stability of moral behavior to Epstein's examination of behavior stability and the stability of linguistic behavior in our samples, we provide a strong test of the existence of moral stability.

Given the two dominant perspectives on the existence of stable individual differences, we examine two different predictions for the degree to which moral behaviors are stable. On the one hand, moral demands may vary wildly across situations (occasions for charity may be fairly infrequent) and these demands may be, well, demanding (charitable donations cost money). If so, moral behaviors may be less stable than behaviors that do not bear moral significance. On the other hand, people care a lot about their moral identity and being seen as moral; the person who sees herself—and presents herself—as honest may experience pressures to be honest from both herself and others. Thus, if people attach special significance to behaving morally, nominally dissimilar situations (e.g., lecturing in class and going out with friends) might be psychologically similar (i.e., seen as affording the opportunity to portray a self-image of integrity; Shoda, Mischel, & Wright, 1994). According to this second view, we would expect higher stability for moral behaviors.

We used data from three samples with a combined total of 186 people, including women with breast cancer (Sample 1; N = 52) and their cohabitating partners (Sample 1; N = 50; for details see Robbins, López, Weihs, & Mehl, in press), healthy adults (Sample 2; N = 73), and individuals with rheumatoid arthritis (Sample 3; N = 11; for details see Robbins, Mehl, Holleran, & Kasle, 2011). All participants wore the EAR (Mehl & Robbins, 2012), which recorded them for brief intervals (30–50 seconds) every few minutes (9–18 min). Sample 1 participants wore the EAR over one weekend, and Samples 2 and 3 participants wore the EAR on two weekends one month apart (Sample 2) or two months apart (Sample 3). In all three samples, participants did not know our specific behaviors of interest, decreasing the likelihood that they would specifically attempt to adjust or regulate their moral behaviors when wearing the EAR.

Trained research assistants coded the EAR files. Each file was coded by three independent coders. In Samples 2 and 3, the files for the first and second weekend were coded by the same coders for any given participant. For all samples, coders listened to all sound files in which participants were talking. Coders then coded each captured conversation (i.e., any file where the participant was recorded talking) for a set of moral behaviors (such as showing gratitude,

apologizing, and showing sympathy or concern) and immoral behaviors (bragging, blaming, being arrogant or condescending). For each sound file, coders rated whether a behavior was present by rating a one if it was. All behaviors were coded for in each of the sound files in which the participants were talking, and any one file could contain more than one kind of moral behavior. This behavior-counting approach to EAR coding complements existing behavioral observation studies that are based on behavior rating systems (Funder, Furr, & Colvin, 2000), and yields data that are based on non-arbitrary, naturally meaningful, and intuitively interpretable metrics (i.e., the percentage of conversations in which a behavior was present, for example, the percentage of a participant's conversations over the course of the EAR monitoring in which the participant expressed gratitude or showed affection), thereby facilitating effect size interpretation (Blanton & Jaccard, 2006; Mehl, 2009; Sechrest, McKnight, & McKnight, 1996).

All EAR files were also transcribed, following guidelines for the Linguistic Inquiry Word Count (LIWC) text analysis program (Pennebaker, Francis, & Booth, 2007). These transcripts were then submitted to the LIWC, which provided us with participants' word usage for a variety of pre-established categories. Specifically, we were given the percentage of words (out of total words spoken) in which participants used particular categories of words (e.g., prepositions). We were specifically interested in neutral categories, and as such, only focused on articles, prepositions, adverbs, space, time, and numbers in our analyses.

Once our data were prepared, we first examined momentary stability. Again, this measurement captured the cohesiveness of one's moral behaviors and language variables by correlating all of one's odd files with all of one's even files for the time period of the monitored weekend. We found that moral behaviors showed high momentary stability—that is, the typical behavior in odd-numbered sound files correlated strongly with the typical behavior in even-numbered sound files (average r across the 14 behaviors = .37). This stability estimate was similar for our language variables (average r = .28; for more details see Bollich et al., under review).

Next, we examined the rank-order stability of moral behaviors and language variables. As a reminder, this test correlated all of one weekend's codings with those from a following weekend, one month, or two months later. These rank-order correlations then demonstrated whether someone who behaves morally and uses particular language variables at one time, is likely to at a later time. Over one month, we found that moral behaviors showed high rank-order stability (average r across all behaviors = .47), as did our language variables (average r = .26). Similarly, over two months, we found that moral behaviors showed high rank-order stability (average r across all behaviors = .52), and our language variables showed moderate rank-order stability (average r = .45).

Taken together, our studies showed that moral behaviors are substantially temporally stable, which provides one kind of evidence for the coherence of moral character. That is, we found that there are individual differences in how morally people behave on a daily basis that are stable from day to day and even from month to month. Moreover, the stability of moral behavior is comparable to the stability of linguistic behaviors. As outlined in detail above, the stability of language behavior, particularly of function words (i.e., words that carry little lexical meaning in a sentence, such as articles and prepositions), constitutes an important benchmark to compare the stability of moral behavior against. Language behavior is an ideal comparison with our moral behaviors for several reasons. For one, it can be extracted with very high reliability using computerized text analysis, which excludes unreliability as a methodological source of instability (Mehl & Pennebaker, 2003). In addition, language behavior can be matched in base rate, and thereby variability, to the moral behaviors extracted from the EAR, which means statistical effects (such as restriction of range) can be ruled out as artifactual causes of instability. Finally, we can select evaluatively neutral language behaviors (e.g., articles or prepositions which are neither considered "desirable" or "undesirable" to use but rather are considered a basic syntactic requirement), which optimizes the comparison to the evaluatively loaded moral behaviors. Pennebaker and King (1999) have established that language variables (such as the use of function words) evidence sufficient stability to warrant construing between-person differences in their uses as reflecting stable individual differences (or, personality).

These findings suggest that a person who behaves in a particularly compassionate or helpful way one day is also likely to be more compassionate and helpful in a similar circumstance than the average person the next day or a month later. This is consistent with other work in personality psychology showing that non-moral behaviors are also relatively stable over time. This is useful because it means that past behavior can be used to predict future moral behavior. In the language of the person-situation debate reviewed earlier, these results provide evidence for stability (exhibiting the same behavior over time), but do not speak directly to the issue of consistency (similarity in trait-related behaviors across different situations). But since stable behavior is evidence of coherence in personality, the results do speak to the issue of coherence.

This research also makes an important methodological contribution to the field, by demonstrating that everyday moral behaviors can be assessed objectively and unobtrusively with the EAR. Given that there are likely biases in both self-reports and other-reports of moral behavior, this technique provides an opportunity to supplement traditional methods. The EAR can be used to examine either the accuracy of self- and other-reports of moral behavior (Bollich, Carlson, Mehl, & Vazire, 2012) or the association between moral behaviors and self-views (while avoiding method overlap).

There are also important limitations of this study. First, the EAR allows us to examine behavior in people's naturalistic environments, but it is limited to detecting acoustic behaviors, and can only record useful data for relatively common behaviors. Behaviors that have no audible component (e.g., cheating) or that are rare (e.g., courageous acts) cannot be effectively studied by means of the EAR. Another limitation of the EAR is that sometimes it is unclear in what context a behavior is performed. The 30- to 50-second clips do provide enough context to interpret most behaviors, but some more complex behaviors may be misinterpreted or missed entirely. Another limitation of our study is that our time interval for stability estimates is rather short (one or two months). Ideally, it would be nice to know whether moral behavior is stable from one year to the next, or even from one decade to the next. Further, these results cannot speak directly to the cross-situational consistency of moral behavior. Our stability results imply that people are consistent even across somewhat disparate situations (i.e., people's situations are probably not identical from one day to the next or from one month to the next), but there is likely also substantial stability in people's situations and contexts, which might account for some (or possibly most) of the stability in moral behavior. To truly test the cross-situational consistency of moral behavior, researchers would need to systematically vary the situations in which participants find themselves, or statistically control for the stability of situations people naturally find themselves in over time. Finally, the EAR methodology objectively documents the presence (or absence) of a moral behavior but provides little information on the motivation that was driving it (or the attribution that was following it)—an important distinction between a vice and a virtue from the Aristotelian tradition. For example, it is possible that an act of displayed affection or compassion is motivated by or instrumentalized for a broader immoral goal. In the same vein, it is also possible that a person comes across as arrogant or bragging in the absence of a malicious intent or even with a distinctly benevolent intent. Combining momentary observational assessments (such as the EAR) with momentary self-report assessments (such as experience sampling) would allow us to obtain "both sides of the story," that is, both the actor's perspective of how an action was intended and the observer's perspective of what it afforded. In this way, researchers might develop a fuller understanding of coherence, based on both nominal and psychological situations.

Character, Large and Small

How exactly does this research relate to the debates with which we began? In a sense, the answer is easy: we provide evidence that moral character exhibits considerable coherence, as manifested by "nominal" stability in behavior.

Although the research is in early stages, it looks like people exhibit substantial stability in moral behavior, and there are individual differences in this stability. Moral behavior exhibits order, and this order can tell us something about what kind of moral agent different people are.

While we've celebrated the peaceful times in which we live, those whose intellectually formative years were spent fighting on one side or the other of the person-situation divide (including some of the present authors) will wonder if either side—preferably their side—gets to gloat. Our answer—at least in our better moments—is neither. These are days of peace, and gloating is unseemly.

The demonstration of coherence in moral character should be taken as good news by personality psychologists and virtue ethicists alike; moral behavior exhibits a certain order, and this order is reflected in stable individual differences. Thus, any situationist who has ever been tempted to say "there is no such thing as character" or "character doesn't exist" has reason to rue the day. Of course, these extreme views are more often attributed to skeptics and situationists by their critics (Kenrick & Funder, 1988) than uttered by situationists themselves, and indeed, some of the putatively offending parties have been at pains to observe that they never said any such thing (Doris 2005: 667). Moreover, the sort of coherence we have identified is driven by stability, not (directly) by consistency, and as such it is something many situationists would already have been prepared to concede, since they have tended to allow the existence of considerable behavioral stability (Doris 2002: 24–26).

If the existence of some coherence in moral character is something all parties can agree to, it remains to be seen what implications our research has for *virtue*. For virtue, it might be thought, concerns matters of importance. Aristotle's paradigm of virtue, the magnanimous man, "honors few things" (1984: 1124b8); it is his nature "to be sluggish and to hold back except where great honor as a great result is at stake," being a man "of few deeds, but of great and notable ones" (1124b24–25). For such a man, one imagines, the kind of behaviors we've identified, such as demonstrating gratitude or apologizing, are likely to be of little concern. Moreover, while it is not obvious that the magnanimous man is a paradigm of moral virtue, or even whether Greek ethics was animated by a notion of morality like that of the post-Enlightenment West (Williams 1993), one doesn't have to be so exalted a personage as he to wonder what we're fussing about. Moral virtue, it might be said, involves things that are *hard*, things that involve *sacrifice*, and those are, manifestly, not the sort of things we captured on the EAR.

The first response to this is simply to acknowledge a methodological limitation we've already noted: "high stakes" behaviors simply occur too infrequently to be the object of the sort of systematic analysis we've attempted. We are just not in a position, empirically, to say anything about such behavior. Virtue ethicists, especially those tempted by "the unity of the virtues"

(Doris 2002: 20–22), might be tempted to extrapolate from the small to the large: maybe the person who comports herself well on minor matters is more likely to do so on major matters. Could be, but our research does not speak to the point, and anyone who is impressed by the limits of behavioral consistency will likely resist the inference.

Our favored response is philosophical: size matters, but it's not the only thing that matters. Great deeds are, true enough, great. But it's no less true that good deeds are good. Aristotle himself recognized that virtues come in different sizes; he distinguished "liberality," which is something like what we would call "generosity," from "magnificence," which is something like what we would call "philanthropy" (1984: 1119b22–1123a33). The appropriate expenditure for a child's gift is less that of the appropriate expenditure for a new hospital wing, but there can be virtue in both expenditures. And Aristotle also recognized lesser virtues, such as ready wit or tact (1127b34–1128a8). Exercises of tact may be less impressive than feats of courage, but they can be impressive nonetheless.

For those fortunate enough to not live in interesting times, the occasions for heroic deeds will be few. This means, first, that for most of us, there will be insufficient data to make confident assessments of stability and consistency with regard to these behaviors. (It would take a long and busy life to contain sufficiently iterated trials of rescuing-a-drowning-child situations—unless one happened to be a lifeguard!). But second, and perhaps more interesting, this means that the moral quality of our life is written in the lower case: engaging in small acts of kindness and gratitude, and refraining from small acts of cruelty and selfishness. As to whether behaviors of this kind are evidence for coherence of character, our research has something to say. And we think it is something interesting.[4]

References

Adams, R. M. (2006). *A theory of virtue: Excellence in being for the good*. Oxford: Clarendon Press.

Albright, L., & Malloy, T. E. (1999). Self-observation of social behavior and metaperception. *Journal of Personality and Social Psychology*, *77*(4), 726–734. doi:10.1037/0022-3514.77.4.726.

[4] We are grateful for helpful comments from Will Fleeson, Mike Furr, and Christian Miller on previous drafts of this chapter. We are further grateful for the feedback we received on this project at the 2012 and 2013 Character Project conferences. This publication was made possible through the support of a grant from The Character Project at Wake Forest University and the John Templeton Foundation. The opinion expressed in this publication are those of the authors and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. The data collection was also supported by NIH grants 5R01AT004698, 3R01AT004698-01A1S1 and 1R03CA137975.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*(6), 1621.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*,1–48. doi:10.1080/10463280802613866.

Annas, J. (2005). Comments on John Doris's lack of character. *Philosophy and Phenomenological Research*, *71*(3), 636–642.

Aristotle. 1984. Translation in Ross, W. D., Urmson, J. O., & Barnes, J. The complete works of Aristotle. Princeton, NJ: Princeton University Press.

Barker, R. G., & Wright, H. S. (1951). *One boy's day: A specimen record of behavior*. Oxford England: Harper.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics redux. *American Psychologist*, *61*(1), 62–71. doi:10.1037/0003-066X.61.1.62.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?. *Perspectives on Psychological Science*, *2*(4), 396–403. doi:10.1111/j.1745-6916.2007.00051.x.

Bollich, K. L., Carlson, E. N., Mehl, M. R., & Vazire, S. (2012). Judging moral character: Self-other agreement on moral traits and behaviors. Talk presented at the 13th annual meeting of the Society of Personality and Social Psychology. San Diego, CA. January.

Bollich, K. L., Doris, J. M., Vazire, S., Raison, C. L., Jackson, J. J., & Mehl, M. R. (under review). Eavesdropping on character: Testing the temporal stability of naturalistically observed moral daily behavior.

Cameron, C., Payne, B., & Doris, J. M. (2013). Morality in high definition: Emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of Experimental Social Psychology*, *49*(4), 719–725. doi:10.1016/j.jesp.2013.02.014.

Campos, B., Graesch, A., Ochs, E., Repetti, R., & Bradbury, T. (2009). Opportunity for interaction? A naturalistic observation study of dual-earner families after work and school. *Journal of Family Psychology*, *23*, 798–807.

College Board. (1976–1977). *Student descriptive questionnaire*. Princeton, NJ: Educational Testing Service.

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass*, *3*, 292–313.

Craik, K. H. (2000). The lived day of an individual: A person–environment perspective. In W. Walsh, K. H. Craik, R. H. Price (Eds.), *Person–environment psychology: New directions and perspectives*, 2nd ed., (pp. 233–266). Mahwah, NJ: Lawrence Erlbaum.

Donnellan, M. B. Lucas, R. E. & Fleeson, W. (2009). Introduction to Personality and Assessment at age 40: Reflections on the legacy of the person-situation debate and the future of person-situation integration. *Journal of Research in Personality*, *43*, 117–119.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press

Doris, J. M. (2005). Replies: Evidence and sensibility. *Philosophy and Phenomenological Research*, *71*(3), 656–677.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*(7), 1097–1126. doi:10.1037/0022-3514.37.7.1097.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist*, *35*(9), 790–806. doi:10.1037/0003-066X.35.9.790.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, *51*, 360–392.

Fleeson, W. (2012). Perspectives on the person: Rapid growth and opportunities for integration. In K. Deaux and M. Snyder, (Eds.), *The Oxford Handbook of Personality and Social Psychology* (pp. 33–63). New York, NY: Oxford University Press.

Fleeson, W., & Noftle, E. (2008). The end of the person-situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass*, *2*, 1667–1684.

Fraley, R., & Shaver, P. R. (1998). Airport separations: A naturalistic study of adult attachment dynamics in separating couples. *Journal of Personality and Social Psychology*, *75*(5), 1198–1212. doi:10.1037/0022-3514.75.5.1198.

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality*, *43*(2), 120–126. doi:10.1016/j.jrp.2008.12.041.

Funder, D. C., Furr, R., & Colvin, C. (2000). The Riverside Behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality*, *68*(3), 451–489. doi:10.1111/1467-6494.00103.

Furr, R. M. (2009a). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369–401. doi:10.1002/per.724.

Furr, R. M. (2009b). Profile analysis in person-situation integration. *Journal of Research in Personality*, *43*(2), 196–207. doi:10.1016/j.jrp.2008.08.002.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, *74*(5), 1337–1349. doi: 10.1037/0022-3514.74.5.1337

Hasler, B. P., Mehl, M. R., Bootzin, R. R., & Vazire, S. (2008). Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality*, *42*(6), 1537–1546. doi:10.1016/j.jrp.2008.07.012.

Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a narcissist: Behavioral manifestations of narcissism in everyday life. *Journal of Research in Personality*, *44*(4), 478–484. doi:10.1016/j.jrp.2010.06.001.

Holzman, P., & Kagan, J. (1995). Whither or wither personality research. In P. E. Shrout, S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 3–11). Hillsdale, NJ: Lawrence Erlbaum.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*(4), 521–551. doi:10.1111/j.1467-6494.1993.tb00781.x.

Leikas, S., Lönnqvist, J., & Verkasalo, M. (2012). Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal situations. *Journal of Personality and Social Psychology*, *103*(6), 1007–1022. doi:10.1037/a0030385.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*(1), 23–34. doi:10.1037/0003-066X.43.1.23.

Kwan, V. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, *111*(1), 94–110. doi:10.1037/0033-295X.111.1.94.

McClelland, D. C. (1985). How motives, skills, and values determine what people do. *American Psychologist*, *40*(7), 812–825. doi:10.1037/0003-066X.40.7.812.

Mehl, M. R. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment*, *18*(3), 340–345. doi:10.1037/1040-3590.18.3.340.

Mehl, M. R. (2009). Naturalistic observation of daily behaviour in personality psychology. *European Journal of Personality*, *23*(5), 414–416.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*(5), 862–877. doi:10.1037/0022-3514.90.5.862.

Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR). *European Journal of Psychological Assessment*, *23*(4), 248–257. doi:10.1027/1015-5759.23.4.248.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*, 857–870.

Mehl, M. R., Pennebaker, J. W., Crow, D., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorded (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments & Computers*, *33*(4), 517–523. doi:10.3758/BF03195410.

Mehl, M. R., & Robbins, M. L. (2012). Naturalistic observation sampling: The Electronically Activated Recorder (EAR). In M. R. Mehl, T. S. Conner (Eds.) *Handbook of research methods for studying daily life* (pp. 176–192). New York, NY: Guilford Press.

Mehl, M. R., Robbins, M. L., & Deters, F. (2012). Naturalistic observation of health-relevant social processes: The electronically activated recorder methodology in psychosomatics. *Psychosomatic Medicine*, *74*(4), 410–417. doi:10.1097/PSY.0b013e3182545470.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, *317*(5834), doi:10.1126/science.1139940.

Messick, D. M., Bloom, S., Boldizar, J. P., & Samuelson, C. D. (1985). Why we are fairer than others. *Journal of Experimental Social Psychology*, *21*(5), 480–500. doi:10.1016/0022-1031(85)90031-9.

Miller, C. B. (2014). *Character and moral psychology*. New York, NY: Oxford University Press.

Mischel, W. (1968). *Personality and assessment*. Hoboken, NJ: John Wiley & Sons.

Mischel, W. (2009). From personality and assessment (1968) to personality science, 2009. *Journal of Research in Personality*, *43*, 282–290. doi:http://dx.doi.org/10.1016/j.jrp.2008.12.037.

Nado, J., Kelly, D., & Stich, S. (2009). Moral judgment. *The Routledge companion to philosophy of psychology.* Abingdon, Engl.: Routledge.

Ozer, D. J. (1986). *Consistency in personality. A Methodological Framework.* Berlin: Springer Verlag.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count: LIWC 2007*. Austin, TX: LIWC. http://www.liwc.net.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. doi:10.1037/0022-3514.77.6.1296.

Pervin, L. A. (1996). Personality: A view of the future based on a look at the past. *Journal of Research in Personality*, *30*(3), 309–318.

Ramírez-Esparza, N., Mehl, M. R., Álvarez-Bermúdez, J., & Pennebaker, J. W. (2009). Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality*, *43*(1), 1–7. doi:10.1016/j.jrp.2008.09.002.

Robbins, M. L., López, A. M., Weihs, K. L., & Mehl, M. R. (in press). Cancer conversations in context: Naturalistic observation of couples coping with breast cancer. *Journal of Family Psychology*, *28*: 380-390

Robbins, M. L., Mehl, M. R., Holleran, S. E., & Kasle, S. (2011). Naturalistically observed sighing and depression in rheumatoid arthritis patients: A preliminary study. *Health Psychology*, *30*, 129–133.

Ross, L. & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, *5*, 2–14.

Sechrest, L., McKnight, P., & McKnight, K. (1996). Calibration of measures for psychotherapy outcome studies. *American Psychologist*, *51*(10), 1065–1071. doi:10.1037/0003-066X.51.10.1065.

Sedikides, C., & Alicke, M. D. (2012). Self-enhancement and self-protection motives. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 303–322). New York, NY: Oxford University Press.

Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014). Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology*, *53*(2), 396-403.

Sherman, R. A., Nave, C. S., & Funder, D. C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, *99*(2), 330–343. doi:10.1037/a0019796.

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, *67*(4), 674–687. doi:10.1037/0022-3514.67.4.674.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers?. *Acta Psychologica*, *47*(2), 143-148.

Swanton, C. (2003). *Virtue ethics: A pluralistic view*. Oxford: Clarendon Press

Tracy, J. L., Robins, R. W., & Sherman, J. W. (2009). The practice of psychological science: Searching for Cronbach's two streams in social-personality psychology. *Journal of Personality and Social Psychology*, *96*(6), 1206.

Tudge, J., Hogan, D., & Etz, K. (1999). Using naturalistic observations as a window into children's everyday lives: An ecological approach. *Contemporary Perspectives on Family Research*, *1*, 109–132.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. doi:10.1037/a0017908.

Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, *4*(8), 605–620. doi:10.1111/j.1751-9004.2010.00280.x.

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, *20*(2), 104–108. doi:10.1177/0963721411402478.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*(5), 1202–1216. doi:10.1037/a0013314.

Williams, B. (1993). *Shame and necessity.*Sather classical lectures, bk. 57. Berkeley: University of California Press.

30 }

Taking Evil into the Lab: Exploring the Frontiers of Morality and Individual Differences

David Gallardo-Pujol, Elizabet Orekhova, Verónica Benet-Martínez, and Mel Slater

Recently, Philip Zimbard, in his book *The Lucifer Effect* (2008), raised the issue of how ordinary people can commit acts of unspeakable cruelty. The basic premise of the book, which received great attention in both the popular media and academia, is that some situations can become so powerful that individual differences among people are extinguished in response to the overwhelming power of the situation. These claims were based on a series of famous experiments that changed the history of psychology. From the fifties through the seventies, a group of impressive scientists pioneered our understanding of the limits of social influence. Asch, Milgram, and Zimbardo are likely the most famous representatives of this effort.

Drawn from the study of classical social psychology situations, the work we present in this chapter aims to challenge some misconceptions and answer important questions. First, we briefly review the person-situation debate with respect to obedience to authority and how the interpretation of Milgram's research has changed over the years. We focus on obedience to authority because of the impact that this program of research had on society, and still has. It was one of the most cited experiments in the twentieth century, and it sparked a flurry of literature (both scientific and non-scientific). Some other reasons are that it was consistently replicated, and that the effect, although not well explained, is consistent, while other experiments have not consistently been replicated, such as Zimbardo's Stanford prison experiment. Second, we present some contemporary re-creations of Milgram's work, from partial replications to virtual-reality simulations; we also present different technologies that can be used to shed light on this debate using modern, updated methods.

Third, based on virtual reality, we summarize the results of an empirical test of the role of individual differences in predicting obedience to authority, or, in other words, whether there are individual characteristics that may explain immoral behavior in Milgram's situation. In this section, we introduce the methods and main results of our research. We chose to use the Big Five framework because it provides a common language and sets a common ground from which further research can stem. Finally, we discuss some of the implications that our findings have regarding the concept of character and propose future directions of research.

1. The Power of the Situation and the Legacy of Milgram

Situations exert powerful influences on human behavior. In Asch's famous experiment, participants were asked to choose which of three lines was the same length as a test line. A third of the subjects gave an obviously incorrect answer after other participants (who were confederates) had also given the incorrect answer. Stanley Milgram conducted a series of obedience experiments in which subjects were told they were participating in a memory test. They were to read a list of word pairs and to administer electric shocks of increasing power to the learner when he gave a wrong answer. The electric shock voltage ranged from 15 to 450 volts. When asked before the experiment, most psychiatrists and psychologists presumed that most subjects would never go above 150 volts (Milgram, 1974). However, in experiments one to four, the majority of participants (60% on average across conditions), regardless of their gender, age or ethnicity, administered up to the maximum voltage, which was clearly marked as highly dangerous (Burger, 2009; Milgram, 1974).

The results of similar experiments conducted later, such as the Good Samaritan experiment by Darley and Batson (1973), pointed in the same direction. In this experiment, participants were exposed to a victim looking for help, and the researchers manipulated the amount of "hurry" participants felt. The degree of help that participants offered to victims varied with the extent to which they had been hurried, showing that the more hurry, the less help they gave. The Stanford prison experiment is often put forward as an example of the situational approach (Haney, Banks, & Zimbardo, 1973). In this experiment, Zimbardo aimed to explore how personality traits could influence cruelty in the prison, but it turned out that roles were more important. The experiment, however, has been criticized for several reasons, ranging from ethical issues to validity and selection biases. Finally, there is recent evidence supporting the importance of environmental factors in determining moral and civic behavior. Keizer and his colleagues showed that manipulating the physical environment, such as covering an alley's wall with graffiti, positively predicts negative

behaviors, such littering, which can be regarded as an example of mildly uncivic and immoral behavior (Keizer, Lindenberg, & Steg, 2008).

The cases we have presented are often taken as examples of moral or immoral behavior (especially the Good Samaritan experiment and Milgram's and Zimbardo's experiments). This has led some authors, especially Doris (2002) and Harman (2009), to claim that there are no cross-situationally consistent and stable meaningful personal characteristics related to morality or moral-like behavior, given the strong influences of situational variables on specific behaviors. Furthermore, Zimbardo (2008) recently stated that "a large body of evidence in social psychology supports the concept that situational power triumphs over individual power in given contexts" (p. x) and that "individual differences should be celebrated, but in the face of strong, common situational forces, individual differences shrink and are compressed" (p. 447).

These claims however are in sharp contrast with the overwhelming evidence gathered by the field of personality psychology, which has clearly demonstrated that personality dispositions have a cumulative impact on a wide range of individual, interpersonal, and social/institutional life outcomes (Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

Personality dispositions have been described using five broad dimensions or factors: extraversion, emotional stability, openness to experience, agreeableness, and conscientiousness (Goldberg, 1993). In lay terms, extraverts (relative to introverts) are more assertive and tend to have a large number of social relationships. "Emotional stability" refers to the intensity and volatility of the emotions we experience on a daily basis. Emotionally stable people tend to perceive the world as a manageable and auspicious place to be, while unstable individuals perceive it as worrying and threatening. In lay terms, "openness to experience" captures intellectual curiosity and preference for variety. "Agreeableness" is related to the quality of social relationships. Agreeable people are usually considerate, generous, kind, trusting, and willing to help. Finally, "conscientiousness" is related to self-discipline, dutifulness, and achievement, among other characteristics.

Personality influences important life outcomes, such as happiness, subjective well-being, and virtue (Ozer & Benet-Martinez, 2006). Gratitude is associated with extraversion and agreeableness (McCullough, Emmons, & Tsang, 2002) and forgiveness is related to openness and agreeableness (Thompson et al., 2005). In the light of these findings, we might think that there is room for some role for personality dispositions to influence morality.

Moral behavior does not depend only on context and situation, and even the leaders of the studies that underscored the power of the situation acknowledge that individual differences in social conformity and obedience to authority exist. While Asch (1952) claimed that individual immunity to group pressures was a function of the person's relationship to himself and others, as independence requires a degree of self-assertion, Milgram (1974) stated that

he was "certain that there is a complex personality basis to obedience and disobedience. But I know we have not found it".

Although Milgram himself advocated the consideration of individual personality variables in understanding disobedience, this idea has historically been dismissed. This is an issue that must be highlighted. In Appendix II of his famous book *Obedience to Authority*, Milgram presented some thoughts that are worth reproducing:

> I had also collected background information on subjects immediately after participation in the experiment. The findings, although generally weak, pointed in the following directions. Republicans and Democrats were not significantly different in obedience levels; Catholics were more obedient than Jews or Protestants. The better educated were more defiant than the less well educated. Those in the moral professions of law, medicine and teaching showed greater defiance than those in the more technical professions, such as engineering and physical science. The longer one's military service, the more obedience—except that former officers were less obedient than those who served only as enlisted men, regardless of length of service. (Milgram, 1974, p. 205)

The variables mentioned in this paragraph are individual characteristics and roles, variables that are heavily influenced by individual cognitive, motivational, and affective dispositions (i.e., personality; Ozer & Benet-Martínez, 2006; Sibley & Duckitt, 2009). In this sense, Milgram was among the first to contemplate psychological individual differences in obedience from a scientific point of view.

Hence, the obedience rates obtained by Milgram can be explained in one of four ways (Funder & Fast, 2010): first, the situational forces toward obedience were stronger than the situational forces toward compassion; second, the participants' dispositions toward obedience were stronger than participants' dispositions toward compassion; third, the situational forces toward obedience were stronger than the dispositional forces toward compassion; and fourth the dispositional forces toward obedience were stronger than the situational forces toward compassion. The third explanation is the one commonly found in textbooks, and the fourth explanation reverses the fundamental attribution error (Funder & Fast, 2010), which attributes human behavior solely to external or situational forces. Interestingly, Funder and Fast (2010) note that explanations three and four are equivalent, but incoherent, while explanations one and two are coherent and also equivalent. Explanations three and four are incoherent because they pit the person against the situation, assuming that only one of these two factors affects actual behavior (Funder & Fast, 2010). The truth is that neither of them by itself can "win" as both take part in the final outcome, which is the extent of obedience or compassion. This point of view, though, has been barely explored, although Milgram suggested that explanations one

and two could be regarded as involving confronting fields of forces, and these fields should operate at the same time.

Milgram's experiment, though, has been often interpreted as a test of inherent opposition or conflict between dispositional forces and situational forces. One possible reason for this interpretation is that Milgram (1974) ended his book with a statement that has prevailed in social psychology over the last fifty years: "Often, it is not so much the kind of person a man is as the kind of situation in which he finds himself that determines how he will act" (p. 205). As a result of this widespread view, very little attention has been paid to the question of whether the observed differences in obedience might reflect individual differences in personality (Blass, 1991).

Specific personality traits, such as authoritarianism, have nevertheless been related to the tendency to be submissive to orders delivered by the moral authorities of the in-group. In fact, Elms and Milgram (1966) were the first to explore the role of this trait with regard to obedience, and found that obedience was positively correlated to authoritarianism on the F-Scale. Later research by Altemeyer (1981, 1988) showed the importance of right-wing authoritarianism for understanding the obedience phenomenon. Right-wing authoritarianism combines a tendency toward submission, aggression, and conventionalism. Trustfulness is another personality trait that has been related to obedience (Holland & Crowne, 1967; Mixon, 1979), although the empirical evidence for this association is mixed. For example, Holland and Crowne (1967) found no evidence of a significant relationship between obedience and interpersonal trust, whereas Miller (1975) reported that trusting participants were more likely to obey than were non-trusting participants. Hostility has also been proposed to explain some of the individual differences in obedience. In an indirect attempt to measure obedience by using a firing paradigm in management settings, Haas (1966) found a positive link between managers' compliance with the order to recommend employees to be fired and their dispositional hostility.

More recent research on obedience suggests that personality traits are related to people's reactions to an obedience-relevant situation, since those who are high in empathic concern express a reluctance to continue with Milgram's experimental procedure earlier than do those who are low on this trait (Burger, 2009). In this study, which we will address later on in the chapter, Burger found that empathic concern and desire for control were related to the number of shocks delivered in a contemporary replication of Milgram's experiment.

2. Contemporary Recreations of Milgram's Experiment

Research on obedience to authority stopped around the mid-1980s as a result of ethical controversies over Milgram's experiments and replications, such as the possible long-lasting effects of the stress that participants had to deal with

during the experiment. This stress could mainly originate from the requests to deliver a potentially harmful shock to a stranger. To Milgram's credit, a year after the experiment he conducted psychiatric interviews, and no long-term distress was found. But many questions remained unanswered, and now there are only a few available options that would satisfy ethical requirements while generating valid results. We review them here.

The two main routes are modification of the original experimental procedure and virtual exposure via videos or virtual reality. The former approach was used by Burger (2009), who carried out the first replication of Milgram's procedure since the formation of ethics committees in the social sciences. Burger replicated Milgram's experiment, but he made participants stop at 150 volts (at the learner's first verbal protest) to ensure that the ethical standards governing his study were met (mainly, not to deliver potentially dangerous shocks). Burger found obedience rates to be only slightly lower than those reported by Milgram for the same condition, 70% versus 82.5%, a difference that was not statistically significant. The choice of the 150-volt threshold was not a trivial decision since a meta-analysis of data from eight of Milgram's experiments revealed systematic disobedience at this point (Packer, 2008). In fact, the more disobedience participants showed at 150 volts, the less they obeyed in the end. These research results show that obedience has not lessened over time, and thus that the social implications of Milgram's work are still relevant (Burger, 2009).

However, there is still controversy over why people refuse to continue. For example, Burger (2009) attributes refusal to give the electrical shock to hearing the victim's protests; whereas, Packer (2008) attributes it to the perception of the learner's right to terminate the experiment. More importantly, if the phenomenon still occurs fifty years later, we may think that the mechanisms underlying obedience and defiance may still be the same. Hence, explaining why a large proportion of participants do not obey is worth studying.

Burger's study is not without limitations. First, it is a partial replication of Milgram's experiment, and it provides no information about what might happen after the 150-volt threshold is exceeded. As such, the study is not fully comparable (Elms, 2009). Second, according to some authors, Burger overscreened participants in order to avoid possible negative reactions to the experiment. That is, if after the initial clinical interview there was any hint of a possible negative reaction during the experiment, that participant was excluded. To Miller (2009), that means that Burger's replication was with a likely biased sample, since less than 50% of the initial sample actually participated, and therefore this hampers any generalizations from Burger's findings (A. G. Miller, 2009). Thus, it remains unclear whether data from the excluded participants would have increased the percentage of disobedient people (Elms, 2009), or whether, perhaps because they lacked the psychological resources needed to resist pressure, obedience among these excluded participants would

have been higher (Masters, 2009). All in all, the psychological underpinnings of Milgram's experiment continue to be subject to reinterpretation (Blass, 1991; Modigliani & Rochat, 1995; Russell, 2011).

The other option for future obedience studies is the use of virtual exposure methods, such as video and TV shows or immersive virtual reality (IVR). Given the ethical constraints on human behavioral research, Slater et al. (2006) introduced the use of IVR. The sense of presence induced in IVR, that is, the tendency for people to behave realistically (Sanchez Vives & Slater, 2005), affords its use for social psychological studies (Blascovich et al., 2002), especially in extreme social situations depicting violence (Rovira, Swapp, Spanlang, & Slater, 2009). With IVR it has become possible to explore behavior in social situations, such as behavior in violent emergencies, without endangering participants' well-being (e.g., Slater et al., 2013).

Slater and his colleagues (2006) fully recreated a Milgram scenario with two different conditions, with and without visual contact with the "virtual" learner. Thus their scenario resembled Milgram's original experiments two and three, which involved contact between the teacher and the learner only by means of voice (experiment two), and where the teacher and learner were in the same room (experiment three). In fact, Slater and colleagues hypothesized that participants' stress would be greater when the learner is present in the same room versus when the learner is hidden (other room). In the former, on the desk in front of the participant was an "electric shock machine" with a shock button, voltage indicators, and a knob for turning up the voltage level. One could think that given the artificiality of the setting, IVR outcomes would radically differ from real-life settings, but participants responded realistically at the subjective, physiological, and behavioral levels (skin conductance analysis, ECG analysis) despite knowing that the situation was not real. In fact, participants showed levels of obedience and disobedience that were similar to those reported by Milgram (Slater et al., 2006).

Interestingly, a re-analysis of the data obtained by Slater et al. (2006) showed that the personality trait of conscientiousness could predict dropping out of the study (Gallardo-Pujol, Muntada-Grimaldi, & Slater, 2011). Conscientiousness is a personality trait capturing individual differences in achievement orientation, order, competence, self-discipline, dutifulness, and deliberation. Conscientious people have a strong sense of dutifulness, but they also adhere scrupulously to their moral precepts and tend to deliberate, which might explain why some would chose to withdraw from Slater's study. This finding, although preliminary, because of the small sample size and power limitations, calls for a more in-depth exploration of the role of individual differences in character with regard to moral behavior in obedience contexts.

Dambrun and Vatine (2010) created an alternative to the immersive virtual environment of Slater et al. This involved an immersive video environment constructed with a video of a real person (an actor) that is programmed to

be shown on a computer screen in a way that makes the experiment seem real time and interactive. Their findings were very similar to Milgram's (1974) original results and the virtual replication results of Slater et al. (2006) with regard to observed levels of obedience, differences in obedience as a function of victim's proximity, and attributions of responsibility. Dambrun and Vatine (2010) also found that both state-anger and right-wing authoritarianism were significant predictors of the maximum voltage shocks administered. They also examined a new variable, the victim's ethnicity, and found that while participants were not significantly more inclined to punish out-group members, they were less anxious and reported a lower level of distress when delivering shocks to them.

Beauvois, Courbet and Oberlé (2012) went a step beyond and transposed the Milgram's paradigm to a TV show in order to test the power of the media, rather than "just" a legitimate authority, such as a scientist. In this setting their findings were very similar to those obtained by Milgram and also by more recent replications, such as Burger (2009), Slater et al., (2006), or Dambrun & Vatine (2010). They used the context of a real TV game show, on French TV, with a live audience and no prizes. This re-creation also involved a confederate learner, an electric-shock generator, and a host (i.e., the show presenter), who acted as a source of authority. Their transposition to the television setting had to satisfy two requirements. The first consisted of staying as close as possible to Milgram's experimental situation while putting the subjects in a credible television context. The second requirement was to make sure that the rapport between the agent of authority and the agent subjected to that authority was very close to the one created by Milgram. In fact, the authors point to the degree of identification with the situation as the main factor for obedience in this context.

This intriguing line of research shows that TV has prescriptive power for influencing people's behavior on a television stage, including counterattitudinal (i.e., behavior that is inconsistent with attitudes), cruel, and immoral behavior (Beauvois et al., 2012). Some time after the show took place, the researchers collected data on personality and political attitudes that can help us understand why 20% of their participants dropped out of the experiment. They found that agreeableness and conscientiousness were positively related to the intensity of the shocks delivered, and also that a left-wing orientation was linked to increased disobedience (Bègue et al., 2014).

Although it could be argued that virtual reality is not the same as reality and, therefore, that the experimental results do not have the same validity, carrying out controlled studies in virtual reality leads to the same degree of ecological validity as compared to traditional lab-based approaches, mainly based on analogues of real-life variables. Virtual reality also provides enough flexibility and reproducibility for the experiment simulations, given that all participants can run through exactly the same script, and the script can be

easily modified to compare groups (Rovira et al., 2009). Individually considered, these characteristics are not unique to virtual reality, but taken together, they make virtual reality simulations an excellent tool to study obedience to authority in our modern days.

In any case, individual differences in obedience have not been studied so far in virtual reality beyond the results we have presented here, and this is what we are going to explore further in this chapter. As we have seen, virtual reality offers a wide array of possibilities to overcome the ethical limitations inherent to Milgram's experiments.

3. The Role of Individual Differences in Obedience to Authority: An Empirical Test

In order to understand the role of individual differences in obedience to authority, we designed a replication of Milgram's original experiment in virtual reality, according to Slater et al.'s (2006) specifications. The script used to recreate this experiment is available upon request. Here, we are going to describe how we tested the influence of individual differences on obedience to authority.

In our approach, we combined an immersive setting (including a virtual avatar in a virtual room) with analog elements, such as a real shock generator and also a real confederate experimenter who was standing by each participant's side (see figure 30.1). The setting that we reproduced was the same that

FIGURE 30.1 View of the experimental setting of our recreation of Milgram's experiment.

Milgram laid out in his experiment number three. In this experimental setting, both the learner and the teacher were in the same room, but with no tactile contact (for obvious reasons in the case of virtual reality). We combined virtual and physical objects to cover the situation with more plausibility. Hence, we used an analog shock generator linked to the computer, and a physical button to deliver the shocks, so the participant interacted with real interfaces. Unlike the experimenter in Milgram's experiment, ours was standing by the participant's side to answer possible questions. However, unlike in the Milgram experiment, this putative authority figure (he was wearing a white lab coat and delivered all the instructions to the participant) never pushed the participants to follow orders, and thus obey, if they expressed hesitation. This was an important and crucial difference between our experiment and Milgram's. While in Milgram's study there was an authority figure in the original experiment pushing the participant to follow orders, in ours such a figure did not exist, since our authority-like figure did not give any orders to follow.

Once we developed the virtual-reality setting, we gathered a sample of eighty males drawn from the general population of Barcelona (Spain), a relatively liberal city in the context of Europe. The reason for choosing only males is that there are sex differences in obedience rates (Milgram, 1974) and thus for the present research we decided, for practical reasons, to block the gender variable. After a careful review of the literature (Gallardo-Pujol, Orekhova, & Benet-Martínez, under submission), we came across a number of individual differences that could influence obedience behavior. We reviewed some of these in the first section of the chapter, such as authoritarianism, right-wing authoritarianism, trustfulness, and hostility (Gallardo-Pujol et al., under submission). However, notice that these individual differences were conceptualized several decades ago and also tap narrow components of broader traits. Thus, we conceptualized their relationship with obedience in terms of modern personality models— namely, the Big Five. Using this broadly accepted taxonomy is important because results in our study can provide a common ground from which specific, comparable hypotheses could be further tested in the future. We developed two hypotheses. On the one hand, the profile of an obedient individual would be characterized by high levels of conscientiousness, agreeableness, emotional stability, introversion, and low openness (i.e., conservatism). On the other hand, the profile of a non-obedient person would fit an extravert who is high in neuroticism and openness to experience, but also low on agreeableness and low on conscientiousness.

Besides broad personality traits, we also collected data on intelligence scores and narrow personality traits, such as aggressiveness. Furthermore, we were deeply interested in understanding how participants perceived situational cues and reacted to them. To do so, we used a new method specifically designed for this aim: the Riverside Situational Q-sort. A Q-sort is a method that requires the participant to sort a set of statements according to his or her

subjective preferences. This taxonomy assesses the existence of psychologically meaningful, important, and consequential situation characteristics that are tied to situation cues, goal affordances, and behavior. These characteristics can be aggregated into eight dimensions—namely, duty, intellect, adversity, mating, positivity, negativity, deception, and sociality (Rauthmann et al., 2014). A few examples of the questions we asked for the experiment situation were: "it is possible to deceive someone," "situation is potentially enjoyable," "situation is potentially anxiety-inducing," or "social interaction is possible." In short, personality, intelligence, and situational cues were the explanatory variables in our study, while we considered the number of shocks and abandoning the experiment (yes/no) as dependent variables.

In our study, we found an obedience rate (i.e., not dropping out of the experiment) of 72%, which was very similar to Milgram's original or subsequent replications (Burger, 2009; Slater et al., 2006). In fact, when we looked at the distribution of the shocks, we found a bimodal distribution with two peaks, at around 150 volts and 400 vo;ts, which coincide with the first victim's complaint and the end of the experiment. Not surprisingly, this distribution much resembles the bimodal distribution found by Packer (2008) in his meta-analysis of all Milgram's experiments. It is worth noting that when we asked the participants to rate how realistically they perceived the situation, using a scale ranging from 1 to 10, most scores were around 7 and 8 (the median score was 6), thus indicating a great degree of immersion.

When we ran the analyses taking into account broad personality data, intelligence, political attitudes, and narrow personality traits, we found that intelligence was the strongest (negative) predictor of the number of shocks. That is, the higher the intelligence, the more likely the participant was to drop out of the experiment. Another significant predictor was conscientiousness, which positively predicted the number of shocks administered. Emotional stability and aggressiveness were also predictive of the number of shocks, so that the more emotionally stable and aggressive the participant, the more shocks he delivered (Gallardo-Pujol, Orekhova, Rodrigues, Benet-Martínez, Slater, in preparation). All these effects were in the small- to medium-size range, around .3 (Cohen, 1988). Finally, we looked at whether the perception of the situation could somehow differentiate between those who dropped out of the experiment and those who kept delivering shocks until the end. Interestingly, we found that those who dropped out of the experiment were less likely to perceive the situation as containing mistrust, deception, lying, betrayal, and hostility. That is, seeing a situation as deceitful is seemingly related to aggression or hostility affordances (Rauthmann et al., 2014). Not surprisingly, all these elements are present in Milgram's experiment to some extent.

We also collected DNA samples to test the possible association between specific genes and obedience to authority. We examined the monoamine oxidase

A (MAOA) gene, since it has been previously linked to aggressive or hostile responses to social stimuli (Gallardo-Pujol et al., 2013), and thus it could to some extent influence the behavior we aimed to explore. We also examined the serotonin transporter gene, which has been previously related to emotional stability and hostile responses as well (Lesch, 1996). Finally, because there is evidence linking empathy, social cognition, and the oxytocin receptor gene (Ebstein et al., 2010), we genotyped two single nucleotide polymorphisms of this gene. As of the time of publication, the results for all these three genes are not available.

4. What's Next? Perspectives on Research on Obedience to Authority

Drawing from previous research, we hypothesized that obedient participants would show a profile of high conscientiousness, high agreeableness, high emotional stability, and low extraversion. We also hypothesized that non-obedient participants would show low emotional stability, high extraversion, low conscientiousness, and high openness to experience. We had no prior hypotheses with respect to intelligence or situational cues.

Our virtual re-creation of Milgram's experiments revealed that personal or individual characteristics (intelligence, personality), as well as the perception of specific situational cues, clearly predict specific outcomes (i.e., the number of shocks) in extreme situations. More explicitly, we found that administering a high number of shocks is predicted by low intelligence, high conscientiousness, high aggressiveness, and high emotional stability. We also found that participants delivering more shocks perceived the situation in Milgram's experiment as more hostile than those who stopped delivering shocks in early stages of it.

Our findings partially support our initial hypotheses, especially with respect to the important role that the personality trait of conscientiousness seems to play in obedience to authority. Hence, conscientious individuals tend to be more obedient and to accomplish their duties. In contrast to preliminary research from our group (Gallardo-Pujol et al., 2011), dutifulness, order, self-discipline, or achievement-striving aspects of conscientiousness seem to prevail over deliberation and adherence to moral standards. In our experimental setting, obedient participants did what they were told, and they did their best to do it well, despite the fact they were delivering potentially harmful shocks. This is in clear harmony with anecdotal observations made by Milgram (1974) in his original experiments when explaining why some participants followed orders, especially when he noted that some proxies for personality variables could be underlying obedience (see the quotation from *Obedience to Authority*, p. 205, in section 1)).

With respect to emotional stability, we found the expected results. High levels of emotional stability are associated with a greater number of shocks delivered. Thus, it seems that individuals showing low emotional stability could be more prone to drop out when they have to face emotional distress while executing the task, such as when the avatar complains about the pain he is feeling. Some researchers have identified that dropping out of the experiment seems to be more linked to personal distress (Cheetham et al., 2009) and emotionally unstable individuals are more vulnerable to stress (Ozer & Benet-Martínez, 2006). Hence, our finding could partially explain why some people do not drop out of the experiment while other people do. The relationship between aggressiveness and inflicting pain on a stranger could reflect underlying dark triad traits (Paulhus & Williams, 2002); or, in lay terms, this would be the profile of a sociopath, although that would require a certain degree of emotional stability, which is the case in our results. Under this view, the teacher could be finding that inflicting pain on a stranger is somewhat reinforcing. Further research is definitely needed to clarify the relative contributions of different profiles to obedience to authority.

Remarkably, another important finding arising from our research is that intelligence was the strongest predictor of obedience. So far, recent approaches to individual differences and obedience to authority have relied only on personality variables (Bègue et al., 2014); our study is the first that directly inquires about the role of intelligence. Again, Milgram's intuitions and anecdotal observations pointed in the same direction that the empirical data do today. He stated that the more educated participants were more defiant than the less educated (Milgram, 1974), and there is an actual relationship between intelligence and educational attainment (Deary et al., 2007). Hence, our findings confirm his observations on educational attainment and obedience.

All in all, immoral behavior, in this particular case, inflicting pain on a stranger, depends on a combination of a set of characteristics, from general dispositions, such as intelligence, to very narrow and specific personality traits, such as aggressiveness. So.it seems that moral behavior might be the result of the combination of general processes and specific ones.

Drawing from our findings, it seems that subjective perceptions of the situation can also exert some influence on our behavior: the relevant situational cues are in the eye of the beholder. Particularly, we found that those participants who dropped out of the experiment perceived the situation with less deception, while those who gave the maximum number of shocks perceived more deception. In this context, deception describes to what extent people perceive a situation to contain mistrust, deception, lying, betrayal, and hostility. Here, deception as we measured it is a newly identified construct, and could be tied to the dark triad traits (i.e., subclinical narcissism, Machiavellianism, psychopathy; Paulhus & Williams, 2002). Deception can be particularly fruitful for research on "darker" aspects of social dilemma games and interpersonal

trust (Rauthmann et al., 2014), and it is consistent with our previous observation about aggressiveness and its relationship to the delivery of shocks.

After rigorous empirical scrutiny, it becomes apparent that personal characteristics do matter in extreme situations, such as obedience to authority, in which people are usually put under strong situational pressures. As we mentioned in the first section of the chapter, the interpretation of Milgram's experiment has evolved over the years, and it is currently often interpreted as a struggle between personal and situational forces. It is, furthermore, often held that the situation is more powerful than the person. We have found evidence that there might be dispositional forces in conflict or situational forces in conflict that operate via perception of situation characteristics, such as the degree of deception that one situation entails.

Additionally, we found some evidence that when a participant had to struggle between continuing to deliver shocks and dropping out of the experiment, there might be different dispositional forces operating in different directions. That is, dispositional forces toward obedience (i.e., high conscientiousness or low intelligence) battled with dispositional forces toward compassion (i.e., emotional stability or high intelligence). Situational forces might operate in the same way and synergistically with dispositional forces. Previous research showed that distance was a well-known situational feature from the original experiments, but we currently know that the perception of hostility in the situation can also push the participant to deliver more shocks.

There are still some issues of debate around our experiment. For instance, there is a key difference between Milgram's original experiment and contemporary replications, since in the former, there was actually someone asking for obedience, while in the latter there is no authority figure. That is, the participants in our experiment, for obvious ethical reasons, had the right to drop out without any questions being asked, at any time during the course of the experiment. While this was also true for participnts in the Milgram experiments, they were prompted up to four times before they could leave the room. However, the "obedience" rates were roughly the same in both experiments. This posits an interesting question about *what* compels participants to deliver a shock in this paradigm. If there is no authority figure pushing participants to shock strangers, we could argue that there might be internal motivations to accomplish the task. The interpretation of our findings should take into account this crucial difference with respect to motivational processes and attributions.

It is still unclear what makes some people (the majority according to Milgram's results) obey at the expense of morality, whereas others do not. As we have attempted to show here, the situation has traditionally been thought to exert a great influence on individual behavior,;yet for more than fifty years, social psychologists and personality researchers have been unable to explain why 36% of Milgram's participants chose to leave before the end of the

experiment. Many have wondered what would happen if we were to go back and investigate more characteristics of the participants who took part in Milgram's experiments. Fortunately, we have developed a working paradigm in virtual reality that could make this possible, and we have paved the way for future research on obedience to authority. Hence, it is worth using modern research methods to tackle, old questions that have remained unchanged for more than half a century.

Although previous research has suggested various interrelationships between personality characteristics and conformity, this work is one of the first to directly tackle this question. In this context, the use of contemporary models and measures of personality (e.g., the Big Five or the HEXACO models) would be helpful because this would create a comprehensive baseline with which to make future cross-study comparisons. Also, for the very first time, we have shown that intelligence might play a key role in obedience.

Understanding the causes of behavior in situations such as Milgram's experiment may well provide significant insights about human nature, and especially about the controversial concept of character. Some authors argue that character as such is hard to study scientifically and cast some doubt on its consistency (Allport, 1937; Doris, 2002; Harman, 2009). However, some researchers have identified a stable, consistent, personological core for the moral domain, based on personality variables such as agency or communion (Walker & Frimer, 2007).

Still, the study of these characteristics in real life is difficult because of the low occurrence of extreme behaviors in which they can be manifested. Approaches like the one we presented in this chapter may offer an excellent framework to reveal these individual differences and set up an ethically controlled, extreme situation in which they may come to the surface of behavior, and may help us to understand how we can prevent obedience to authority when it implies wrongdoing.

Given the prominent role of intelligence as a strong predictor of disobedience or compassion, it is worth exploring the link between moral reasoning and cognitive function. Early research identified intelligence as a crucial aspect for moral reasoning (Kahn, 1983). In adolescents, moral reasoning is associated both with intelligence and social adjustment (Hanks, 1985). More recent research has identified executive functioning as a predictor of moral reasoning, but this effect is moderated by general reasoning (Cottone et al., 2007). Since personality and intelligence are closely related to emotional intelligence as well, it would be an interesting line of research to test whether emotional intelligence would predict obedience in the same way, and if improvements of emotional intelligence, via training programs or skills regulation, could increase compassion when it would be wrong to obey. There is some evidence that this could be the case (Hinnant et al., 2013). Ultimately, the

more intelligent one is, the more critical one is, and therefore the more one questions orders.

In terms of upcoming research, it should include experimental manipulations to replicate the original ones (i.e., to show that situational effects are consistent over time), but also ingroup/outgroup manipulations, gender, ethnicity, and so on. Actually, combining virtual reality with classical social psychology experiments and personality psychological approaches allows us to recreate other milestones, such as the bystander effect (Slater et al., 2013) or the Good Samaritan effect (Darley & Batson, 1973). Virtual simulations should be complemented with alternative techniques, such as lexical approaches to obedience and disobedience (Passini & Morselli, 2010) that should be regarded with interest by personality and social psychologists. In the example we mention, different clusters of words and meanings that are associated with obedience and disobedience are explored.

To conclude, taking all the aforementioned factors into account will also lead to the better understanding and prediction of human behavior in specific situations, especially when we refer to moral behavior. We wholeheartedly believe that the study of individual differences in obedience to authority is a fruitful direction for future research and a promising viewpoint in personality and social psychology.[1]

References

Allport, G. W. (1937). *Personality: a psychological interpretation. Personality: a psychological interpretation*. Oxford England: Holt.

Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg: University of Manitoba Press.

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco, CA: Jossey-Bass.

Asch, S. (1952). *Social Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Beauvois, J-L., Courbet, D., & Oberlé, D. (2012). The prescriptive power of the television host: A transposition of Milgram's obedience paradigm to the context of TV game show. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *62*(3), 111–119. doi:10.1016/j.erap.2012.02.001.

[1] Parts of this manuscript were presented at the Character Project Conferences in 2012 and 2013 and at the 16th European Conference on Personality, Trieste, Italy, July 2012 and also are to appear in Gallardo-Pujol, Orekhova, & Benet-Martínez (2014, submitted). The authors also acknowledge Christian Miller and Mike Furr for their invaluable comments and edits on earlier versions of this manuscript. This chapter was made possible through the support of a grant from The Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of The Character Project, Wake Forest University or the John Templeton Foundation. The equipment used was funded by MS' ERC Advanced Grant TRAVERSE.

Bègue, L., Beauvois, J-L., Courbet, D., Oberlé, D. Lepage, J., & Duke, A. D. (2014). Personality Predicts Obedience in a Milgram Paradigm. *Journal of Personality*. doi: 10.1111/jopy.12104.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J.N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry 13*: 103–124. doi:10.1207/S15327965PLI1302_01.

Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, *60*(3), 398–413. doi:10.1037/0022-3514.60.3.398.

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *The American Psychologist*, *64*(1), 1–11. doi:10.1037/a0010932.

Cheetham, M., Pedroni, A. F., Angus, A., et al. (2009). Virtual Milgram: Empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers of Human Neuroscience*, *3*, ArtID 29. doi:10.3389/neuro.09.029.2009.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Cottone, J., Drucker, P., & Javier, R. (2007). Predictors of Moral Reasoning: Components of Executive Functioning and Aspects of Religiosity. *Journal for the Scientific Study of Religion 46*(1): 37–53.

Dambrun, M., & Vatine, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology*, *40*, 760–773.

Darley, J., & Batson, C. (1973). "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, *27*, 100–108.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. New York: Cambridge University Press.

Ebstein, Richard P; Israel, Salomon; Chew, Soo Hong; Zhong, Songfa; Knafo, Ariel (2010). Genetics of human social behavior. *Neuron* vol. 65 (6) p. 831-44

Elms, A. C. (2009). Obedience lite. *American Psychologist*, *64*(1), 32–36. doi:10.1037/a0014473.

Elms, A. C., & Milgram, S. (1966). Personality characyeristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality*, *1*(4), 282–289.

Funder, D. C., & Fast, L. A. (2010). Personality in Social Psychology. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (5th ed., pp. 668–697). Hoboken, NJ: John Wiley & Sons.

Gallardo-Pujol, D., Andrés-Pueyo, A. & Maydeu-Olivares, A. (2013). MAOA genotype, social exclusion and aggression: an experimental test of a gene–environment interaction. *Genes, Brain and Behavior*, *12*(1), 140–145.

Gallardo-Pujol, D., Muntada-Grimaldi, S., & Slater, M. (2011). Conscientiousness predicts dropping out of the experiment in a virtual reprise of Milgram's paradigm. *Association for Research in Personality*. Riverside, California.

Gallardo-Pujol, D., Orekhova, E., & Benet-Martínez, V. (under submission). Under Pressure to Obey and Conform: When the Power of the Situation is not enough.

Gallardo-Pujol, D. Orekhova, L. Rodrigues, P., Benet-Martínez, V. & Slater, M. (in preparation). Milgram was right: Individual differences predict the number of shocks delivered in Milgram's paradigm.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(January), 26–34.

Haas, K. (1966). Obedience: Submission to destructive orders as related to hostility. *Psychological Reports*, *19*, 32–34.

Haney, C., Banks, W. C., & Zimbardo, P. G. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, 1, 69–97.

Hanks, R. (2006). Moral Reasoning in Adolescents: a Feature of Intelligence or Social Adjustment? *Journal of Moral Education*, *14*(1), 43–55. doi:10.1080/0305724850140106

Harman, G. (2009). Skepticism about character traits. *Journal of Ethics*, *13*, 235–242.

Hinnant J. B., Nelson J. A., O'Brien M., Keane S. P., & Calkins S. D. (2013). The interactive roles of parenting, emotion regulation and executive functioning in moral reasoning during middle childhood. *Cognition and Emotion 27*, 1460–1468.

Holland, C. H., & Crowne, D. P. (1967). *Sources of Variance in the Experimental Investigation of Behavioral Obedience*. Unpublished Doctoral Dissertation, Univ. of Connecticut

Kahn, J. V. (1983). Moral reasoning of Piagetian-matched retarded and nonretarded children and adolescents. *Journal of Genetic Psychology*, *143*(1st Half), 69–77.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, *322*, 1681–1685. doi:10.1126/science.1161405.

Lesch, K.-P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., . . . Murphy, D. L. (1996). Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region. *Science*, *274*(5292), 1527–1531. doi:10.1126/science.274.5292.1527

Masters, K. S. (2009). Milgram, stress research, and the Institutional Review Board. *American Psychologist*, *64*(7), 621–622. doi:10.1037/a0017110.

McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, *82*(1), 112–127. doi:10.1037//0022-3514.82.1.112.

Milgram, S. (1974). *Obedience to Authority: An experimental view. Rain*. London: Tavistock. doi:10.2307/3032069.

Miller, A. G. (2009). Reflections on "Replicating Milgram" (Burger, 2009). *The American Psychologist*, *64*(1), 20–7. doi:10.1037/a0014407.

Miller, F. D. (1975). *An Experimental Study of Obedience to Authorities of Varying Legitimacy*.

Mixon, D (1979) Understanding shocking and puzzling conduct. In Ginsberg, G.P (ed) *Emerging Strategies in Social Psychological Research*, Chichester: Wiley, 155-176.

Modigliani, A., & Rochat, F. (1995). The role of interaction sequences and the timing of resistance in shaping obedience and defiance to authority. *Journal of Social Issues*, *51*(3), 107–123. doi:10.1111/j.1540-4560.1995.tb01337.x.

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421.

Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments: A meta-analytic review. *Perspectives on Psychological Science*, *3*(4), 301–304. doi:10.1111/j.1745-6924.2008.00080.x.

Passini, S., & Morselli, D. (2010). The obedience–disobedience dynamic and the role of responsibility. *Journal of Community & Applied Social Psychology*, *20*, 1–14. doi:10.1002/casp.

Paulhus, D. L., & Williams, K. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*, 556–568.

Rauthmann, J. F., Gallardo-Pujol, D., Nave, C. S., Guillaume, E. M., Jones, A. B., Funder, D. C., & Todd, E. (2014). Major Dimensions of Situation Characteristics The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. doi: 10.1037/a0037250.

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345.

Rovira, A., Swapp, D., Spanlang, B., & Slater, M. (2009). The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioral Neuroscience*, *3*, 59. doi:10.3389/neuro.08.059.2009.

Russell, N. J. C. (2011). Milgram's Obedience to Authority experiments: Origins and early evolution. *British Journal of Social Psychology / British Psychological Society*, *50*(Pt 1), 140–162. doi:10.1348/014466610X492205.

Sanchez--Vives, M. V., Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience 6*, 332–339. doi:10.1038/nrn1651.

Sibley, C. G., & Duckitt, J. (2009). Big Five personality, social worldviews, and ideological attitudes: Further tests of a dual process cognitive-motivational model. *Journal of Social Psychology*, *149*(5), 545–561. doi:10.1080/00224540903232308.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS One*, *1*(1), e39. doi:10.1371/journal.pone.0000039.

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., et al. (2013). Bystander Responses to a Violent Incident in an Immersive Virtual Environment. *PLoS ONE*, *8*(1); e52766. doi:10.1371/journal.pone.0052766.

Thompson, L. Y., Snyder, C. R., Hoffman, L., Michael, S. T., Rasmussen, H. N., Billings, L. S., . . . Roberts, D. E. (2005). Dispositional forgiveness of self, others, and situations. *Journal of Personality*, *73*(2), 313–59. doi:10.1111/j.1467-6494.2005.00311.x.

Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of personality and social psychology*, 93 (5), 845.

Zimbardo, P. (2008). *The Lucifer Effect: Understanding How Good People Turn Evil*. New York: Random House.

31 }

War Crimes: Causes, Excuses, and Blame

Matthew Talbert and Jessica Wolfendale

Introduction

In this chapter we address the following questions: Do war crimes result primarily from failures of character or from situational influences? And, are those who commit war crimes morally responsible for their actions? In response to the first question, we argue that a certain type of situationist account of war crimes is mistaken, and we offer an alternative dispositional account of war crimes. In response to the second question, we argue in favor of a theory of responsibility under which military personnel may be held morally responsible for war crimes in most cases regardless of whether situational forces or dispositions influenced their behavior.

1. Situationism, War Crimes, and Responsibility

A large body of experimental research in social psychology is thought to support situationism:[1] the view that human behavior is explained to a surprising extent by the situational factors to which an agent is exposed, rather than by the agent's pre-existing temporally stable and cross-situationally consistent character traits. One of the striking findings of these experiments is that seemingly minor situational factors appear to disproportionately influence behavior in situations where we would expect even weak character traits to have a much

[1] We recognize that there are several different views that could be described by this term. Here, we use "situationism" to refer to the cluster of philosophical theories that interpret the social science experiments referred to in section 1 as demonstrating that external situational forces explain and determine human behavior to a far greater extent than any pre-existing character traits (Doris 2002; Harman 2000), if such traits exist at all. See Miller (2014), chapter 4, for a helpful review of the different situationist claims in psychology and philosophy.

stronger influence. For example, in Stanley Milgram's famous experiments on destructive obedience, the experimenters were able to persuade normal adults to deliver what they thought were increasingly painful shocks to another person even when that person demanded to be released from the experiment.[2] In another famous experiment, Philip Zimbardo and his colleagues placed willing "normal, healthy American college students" (Haney, Banks, & Zimbardo 1973, 89) in a simulated prison environment housed on the campus of Stanford University. Very shortly after the experiment began, the students role-playing prison guards began to mistreat the students posing as prisoners. Zimbardo concluded that the behavior he observed was "the result of an intrinsically pathological situation which could distort and rechannel the behavior of essentially normal individuals" (1973, 90).

Experiments like those of Milgram and Zimbardo are often thought to tell us something about the commission of war crimes. It is therefore not surprising that a number of philosophers and psychologists (Doris & Murphy 2007; Roth 2004; Snow 2009a; Zimbardo 2004, 2008) have argued that situationism can explain how seemingly ordinary military personnel can become perpetrators of war crimes. According to Zimbardo, a situationist account of war crimes "propels external determinants of behavior to the foreground, well beyond the status as merely extenuating circumstances" (2004, 21). Unlike characterological accounts, situationist accounts of war crimes look to the external situational forces to which individuals are subjected for the explanation of their destructive behavior.

Situationist accounts of war crimes are not only characterized by a focus on the objective features of the situations in which agents act, such as whether a situation exposes agents to extreme stress. Situationist accounts also typically involve a particular conception of the relationship between an agent's mental states, such as beliefs, goals, emotional states (and so forth)[3] and external situational forces. While some situationist accounts acknowledge that agents' mental states can play an important causal role in the commission of war crimes, these mental states are themselves construed as the product of external situational forces, such as exposure to ideology. Thus it is sometimes claimed that because these mental states are imposed upon the agent, the agent's moral responsibility may be called into question, a point we discuss in section 3.

John Doris and Dominic Murphy (2007) defend a situationist account of war crimes that explores the effects of both immediate and temporally extended situational factors on soldiers' behavior. In addition, they explore the implications of their account for soldiers' moral responsibility for war crimes,

[2] The learner was a confederate of the experimenter. For a complete discussion of the experiments, see Milgram (1974).

[3] We intend "mental states" to refer broadly to an agent's psychological states, including cognitive states, such as beliefs and judgments, desires, and emotional states.

arguing that the strength of the situational forces to which soldiers are exposed supports "a strong *general presumption* that perpetrators of wartime atrocities are not morally responsible" (2007, 28; emphasis in the original).

Doris and Murphy interpret the situationist experimental record as indicating, among other things, that confusing, stressful, and ambiguous stimuli—of the sort present in Milgram and Zimbardo's experiments—can "adversely affect moral functioning" (2007, 35). Doris and Murphy argue that if this result occurs in the lab, then there is "good reason to think that the radical confusion of the battlefield will have related effects" (2007, 35). And if, as Doris and Murphy claim, soldiers who commit war crimes "manifest cognitive impairments that profoundly degrade their capacity for moral judgment" (2007, 26), then perhaps they are excused for their behavior. This conclusion follows from the belief—accepted by many philosophers—that an agent is morally responsible for her behavior only if she possesses "normative competence": the capacity to assess and respond to moral reasons.[4]

According to Doris and Murphy, two kinds of situational forces cause cognitive and emotional[5] degradation in soldiers: situational pressures on the battlefield itself, and " 'distal,' or 'wasting' pressures," which are "circumstances that may profoundly affect military personnel, beginning at considerable spatial and temporal distance from the point at which atrocity occurs" (2007, 39). Both these kinds of situational forces, they claim, degrade soldiers' normative competence to the point that they can no longer appropriately exercise their capacity for moral reflection or recognize the moral impermissibility of certain kinds of behavior, factors that can increase the likelihood that they will be involved in the commission of war crimes.[6]

The evidence for the extremely debilitating physical, cognitive, and emotional effects of combat is very strong (Fontana & Rosenheck 1999; Frésard 2004), and so we share Doris and Murphy's view that "the presence of excusing conditions *in combat* is highly probable" (2007, 38).[7] However, Doris and Murphy also claim that soldiers who are involved in non-battlefield war crimes (such as institutionalized torture) should typically be excused for their crimes.

According to Doris and Murphy, the short- and long-term effects of military training and culture shape soldiers' behavior, emotional states,

---

[4] For influential accounts of responsibility that emphasize the capacity to respond to moral reasons, see Wolf (1990); Wallace (1996); and Watson (2004).

[5] Cognitive degradation (as Doris and Murphy use the term) refers to combat's effects on the capacity for rational deliberation among military personnel; whereas psychological degradation refers to the emotionally deleterious impact of combat.

[6] Doris and Murphy use the word "degrade," but this is misleading since at least some of the distal pressures they discuss are better described as obstructing soldiers' exercise of cognitive capacities, rather than as degrading those capacities.

[7] Emphasis added.

and beliefs in ways that compromise their exercise of normative competence (2007, 39). But unlike combat pressures, distal pressures impair soldiers' normative competence by leading them to believe that objectionable behavior is morally permissible or even required. This occurs through a number of interrelated processes, including (but not limited to) the inculcation of obedience to authority, group conformity, desensitization to the use of violence (some soldiers even learn how to "kill prisoners and how to sexually mutilate women" [2007, 39]), dehumanization of the enemy, and exposure to racist and other political ideologies justifying the use of extreme violence against perceived enemies. In summary, the combination of these processes "reconfigures the range of available behavioral options" (2007, 39) for soldiers. As a result, it may no longer be reasonable to expect perpetrators of war crimes who have been exposed to these forces to see their behavior as morally wrong. In Doris and Murphy's view, this outcome arises because distal pressures restrict the scope of soldiers' moral reasoning so that soldiers may cease to possess the ability to come to alternative judgments about how to behave. In the case of war crimes, this may mean that it is unfair to hold perpetrators responsible for their behavior because it is unreasonable to expect them to avoid such behavior since, through no fault of their own, they regard their actions as permissible. We consider the plausibility of this view of responsibility in sections 3 and 4.

2. A Dispositional Approach to War Crimes

One problem for Doris and Murphy's approach is that it fails to account for the variety and degree of perpetrators' enthusiasm for their tasks. It is difficult to explain these differences among perpetrators by reference to situational forces alone, since perpetrators exposed to what appear to be the same situational forces vary greatly in their enthusiasm for engaging in war crimes, and in their attitudes toward their actions (Blass 1993, 34). In addition, Doris and Murphy's account fails to recognize that military culture and training does not merely impose mental states on soldiers from without but cultivates *dispositions* that reflect and are shaped by soldiers' unique individual sets of beliefs, emotional states, and behaviors.

For these reasons we propose an account of war crimes that is sensitive to the insights afforded by the situationist experiments but that emphasizes the way in which war crimes can be conceived of as expressions of soldiers' character and moral agency. War crimes are crimes of character in the sense that perpetrators' character traits play a significant role in explaining why certain kinds of war crimes occur.

By *character traits*, we mean behavior-informing dispositions that are relatively stable over time, and that are sustained by an agent's beliefs, desires, and

values.[8] This conception of character traits is consistent with the social-cognitivist view that personality is best construed as a complex interaction between external situational factors and agents' internal psychological and affective processes. This theory of personality was first proposed by Walter Mischel (1973) and then significantly developed by Mischel and his colleagues over the next 30 years (see Mischel & Shoda 1996), and has gained support from both social psychologists (Ross & Nisbett 2011) and philosophers (Russell 2009; Snow 2009b).[9] According to this theory, it is a mistake to see personality as comprising a set of cross-situationally stable character traits that guide an individual's behavior in a way that is largely independent from the specific situations in which she acts.

Instead, personality should be understood as a cognitive-affective personality system (CAPS): "a dynamic system or organized network of interconnected and inter-acting cognitions and affects" (Mischel & Shoda 1996, 417) that is activated in social settings. In CAPS theory a situation cannot be characterized without understanding how agents in that situation construe the meaning of the situation, and how that construal interacts with their preexisting beliefs, goals, and self-conceptions. In fact, CAPS theory challenges the very legitimacy of the distinction between the person and the situation. As Mischel and Shoda explain, "[T]he concept of 'situation' must go beyond external stimuli to encompass internal, intrapsychic conditions and states in which people's experiences and actions are contextualized" (1996, 422). What the situation *is* for an individual therefore cannot be determined merely from the objective features of the situation but depends crucially on how agents interpret the meaning of that situation. This is in contrast to situationist theories that depict human behavior as resulting from a battle between the force of situations and the power of the person, where the former typically overcomes the latter (Mischel & Shoda 1996, 421). CAPS theory is therefore unique in rejecting the assumption that we can determine the meaning of situations independently from how agents' construe the meaning of those situations.

In CAPS theory, character traits can be conceived of as interconnected "social-cognitive" units involving beliefs, emotions, and judgments that are "activated in response to situational variables . . . or internal stimuli" (Snow 2009b, 31–33).[10] Repeated "activation" of these cognitive-affective units leads to relatively stable behavioral patterns and traits. Nancy Snow illustrates this process through a discussion of how a person may develop the trait of irritability. Like other CAPS traits, the trait of irritability consists of "thoughts, affects, and

[8] Note that this definition does not make cross-situational consistency a requirement of character traits. That said, this definition would still include traditional character traits, such as honesty. For example, a person would possess the trait of honesty under this definition if her disposition to tell the truth informed her behavior over time and was sustained by her beliefs, emotions, and desires (for example, her belief that good people don't lie and her desire to be a good person).

[9] For a critical discussion of the CAPS theory, see Miller (2014), chapter five.

[10] See also Miller (2014), chapter five.

representations of plans, strategies, and values, . . . ready to be activated through external stimuli or even through internal stimuli, such as thoughts or imaginings" (2009a, 32). When the irritable person is exposed to certain stimuli (for example, when someone bumps into her on the street), this activates beliefs ("People are so thoughtless!"), emotions (annoyance), and goals ("I'm going to avoid busy streets") that reinforce the trait as a single unit and can lead the trait to become generalized across objectively different situations, such as crowded movie theaters and public transport. Thus, we must understand the subjective meaning that a situation has for an agent in order to understand that agent's behavior in a given situation.

So CAPS theory can explain both how an agent's behavior across a variety of apparently different situations can arise from a stable individual profile if those situations contain a feature that is "psychologically salient" to the agent (Miller 2014, 117). Similarly, the theory offers an explanation for why an agent's behavior might vary considerably across situations that seem (to the outside observer) very similar if those situations lack such a feature. To illustrate, suppose that, due to a traumatic childhood incident, Marissa is frightened of people in uniforms (police officers, fire fighters, military personnel, and so forth) and consistently reacts to the sight of such individuals by actively avoiding them (walking the other way, for example). To an external observer, situations involving people in uniform might be very different from each other. It is only when we understand the meaning that "people in uniforms" has to Marissa that we can understand the consistency of her actions as arising from a stable disposition related to her unique personal history and psychological makeup.

On our view, CAPS theory provides the most promising framework through which to examine the causes of war crimes. In contrast to Doris and Murphy's view, we suggest that the beliefs and attitudes (and resultant behavior) that are inculcated through military training and culture are not properly thought of as external or alien to individual soldiers. Instead, we argue that military training and culture leads to the development of CAPS traits in individual soldiers by interacting with and shaping their goals, actions, beliefs, and emotions related to their military service and the justification of military force more broadly. At a general level, for example, military training and culture aims to cultivate beliefs (for example, that a particular war is justified), emotions such as group loyalty, and actions (for example, obeying orders) that lead to stable and predictable behavioral patterns in response to particular situations.[11] This process engages soldiers' self-identity, goals, and values so that military service has meaning for them. For example, it is important for military personnel to view their role and the causes for which they fight as morally justified, even

[11] See Dyer (2004), chapters one and two; and Wolfendale (2007), chapter six, for detailed discussions of the aims and impact of basic training in the military.

righteous. To this end, military institutions frequently utilize language, rituals, and practices that present a consistent and constantly reinforced ideal of military service and the role of military personnel as involving the embodiment of moral virtues, including sacrifice, honor, loyalty, self-discipline, and integrity (Wolfendale 2007, 54). This process succeeds if and when individual soldiers successfully internalize these ideals in ways that have significance for them, given their own particular goals and self-conceptions. Thus, while the closed environment of the military means that there is likely to be significant conformity in soldiers' beliefs about how and why they fight, there are still important individual differences between soldiers that reflect the unique ways that soldiers understand the meaning and significance of their military service.

In relation to war crimes, this approach provides important insights into how soldiers can come to view war crimes as permissible and consistent with military values, with being a good soldier, and with their own self-image. Research on the causes of war crimes (Bandura 1999; Frésard 2004) has found that war crimes are likely to occur when the following conditions apply: a specific act or policy, such as a policy of torture, has been authorized and rationalized by military authorities; those executing the policy see their role as requiring obedience; the policy's intended subjects have been dehumanized; the practice has been "sanitized" (for example, referred to in euphemistic terms); the act or policy is treated as legitimate routine behavior consistent with or even required by important military objectives; and the responsibility of those carrying out the practice in question has been displaced onto others. This combination of factors can lead to what psychologist Albert Bandura calls "moral disengagement," where an agent's "moral self-sanctions can be dis-engaged from inhumane conduct" (1994, 194). This process of disengagement is essential for those involved in the commission of war crimes to be able to maintain an intact self-image and self-approval (Bandura 1994, 196).

The factors we have just listed have in common the aim of cultivating specific beliefs, emotions, goals, and self-conceptions in those who will carry out these actions. However, it would be a mistake to interpret the foregoing list as representing a set of ongoing external situational or distal forces that "imprint" soldiers with particular beliefs. Instead, these factors represent the ways in which military training and culture *engage* with the moral agency of military personnel. As noted above, military personnel develop CAPS traits related to their military goals and their specific duties. In the case of war crimes, perpetrators interpret wrongful conduct as permissible and even honorable as a result of the interaction between their own unique goals, values, and beliefs regarding their actions and roles, and exposure to factors such as the rationalizations and justifications offered by military and political leaders.

This process is evident in the testimony and other personal documents of perpetrators of war crimes that reveal how they frequently aim to reconcile their actions with their particular individual self-conceptions, goals,

and values in distinctive and individual ways. For example, in their analysis of the moral psychology of the Holocaust, Herlinde Pauer-Studer and David Velleman (2011) discuss a number of men who participated in the Nazi genocide, and who documented their experiences at the time in private journals and letters.[12] Felix Landau, an Austrian who had been a member of the SS since 1934, served in an *Einsatzcommando* unit that carried out mass killings in occupied Poland. Yet, although he was a long-time member of the SS, he did not see his role as motivated by racial hatred or ideology. Instead, he describes his experience as involving tasks that needed to be done because they were ordered but that he found both unsoldierly and unpleasant. (He writes that he had "little inclination to shoot defenceless people—even if they are only Jews. I would far rather good honest open combat [*sic*]" [Pauer-Studer & Velleman 2011, 345].) He took pride in his hard work, writing, "Everyone admired my work" (347). These and other diary entries reveal that Landau's self-image was tied closely to his perception of himself as a good worker and a good soldier, who did not behave excessively or sadistically in carrying out his orders. In fact, he was critical of overly "enthusiastic" shooters (348). In contrast, Karl Kretschmer, a German pharmacist who served in a unit operating on the Russian Front that carried out mass executions, suffered serious psychological distress at his participation in the killings. However, unlike Landau, Kretschmer interpreted the killings as necessary to protect the German race from the existential threat posed by the Jews ("My comrades are literally fighting for the existence of our people" [Pauer-Studer & Velleman 2011, 349]), and construed his distress as evidence of personal weakness rather than as evidence of any immorality in what he was required to do.

These examples illustrate how two individuals involved in essentially the same war crimes and exposed to similar situational forces in combat and during training interpreted the meaning of their actions in very different ways. These differences affected how and why they acted as they did. For example, Landau's distaste for killing unarmed civilians and his focus on being a good worker meant that he did not pursue the killing of civilians if he was able to avoid it (by, for example, volunteering for other tasks), and when killing was required of him rejected any display of sadism or enjoyment in killing as simply unprofessional. On the other hand, Kretschmer's belief in an existential threat posed by the Jews to the German people allowed him to justify the Nazi's genocidal project as a form of national self-defense. Thus, he saw his personal challenge as overcoming

[12] Pauer-Studer and Velleman use these examples to illustrate how moral values can become inverted through social and other processes, leading to the acceptance of wrongful behavior as consistent with deeply held moral principles.

his revulsion and so he actively sought to become better at his work: "[I]t is a weakness not to be able to stand the sight of dead people; the best way to overcome it is to do it more often" (2011, 349). Unlike Landau, then, Kretschmer's beliefs about his role would have led him to at least attempt to seek out tasks involving killing.

One final example will further illustrate the importance of understanding how perpetrators construe the situations in which they act and their own behavior within those situations. A Brazilian torturer interviewed by Martha Huggins explained his actions as follows: "I don't use . . . violence outside the standard of my conscience as a human being. I'm a conscientious professional. I know what to do and when to do it" (Huggins 2000, 63). This quote suggests that this perpetrator construed his actions as a legitimate exercise of his professional duties and as consistent with his self-conception as a decent man and a good soldier—he only uses violence "on the job"—that is, for legitimate military purposes. This torturer's self-conception no doubt uniquely shaped how he performed his duties, in contrast to, say, a torturer who viewed torture as the infliction of punishment on worthless scum. These three examples illustrate the fact that perpetrators of war crimes are (like all agents) what anthropologist Alexander Hinton calls "meaning makers" who "comprehend and construct meaning out of their social lives" (2004, 28). Ignoring this fact leads to a failure to understand the differences in motivations and attitudes that can explain (at least partially) how and why individuals become perpetrators of war crimes.

The foregoing discussion illustrates the differences between our theory of war crimes and the situationist account offered by Doris and Murphy and others. Our view differs from the situationist theory of war crimes in two important respects. Firstly, in our view, to understand the behavior of military personnel, it is not sufficient to enumerate the different external pressures to which they are exposed. It is also important to understand how individual military personnel themselves construe and interpret the situations they encounter, and what meaning those situations have for them. Secondly, the situationist account offered by Doris and Murphy and others tends to give the impression that the beliefs, goals, emotions, and self-conceptions of military personnel are *imposed* upon them by external sources such as military culture. But this perspective ignores and minimizes the agency of military personnel and downplays the active role that military personnel play in ascribing personal meaning and significance to their actions.

In our account, by contrast, soldiers' behavior can be understood as the result of a constantly evolving and unique interactive relationship between soldiers' individual psychology and personal history and the circumstances in which they act. Thus our view offers both greater explanatory power in relation to war crimes and recognizes military personnel as agents who shape and guide their behavior in unique ways.

3. Excusing Perpetrators

We now consider the moral responsibility of those who commit war crimes. To say that someone is morally responsible for her behavior is to say, in part, that she is an appropriate candidate for various moral responses on account of that behavior. These responses will be positive or negative depending on the quality of the relevant action. Negative responses are associated with a judgment that the person is morally responsible for bad or wrong behavior. Such negative responses—for example, attitudes of resentment or indignation—constitute moral blame. If a person is a proper target of these responses, then she is morally responsible in the sense of being open to moral blame; she is, in a word, blameworthy. Since war crimes are morally objectionable, questions about a perpetrator's moral responsibility largely reduce to the question of whether it is appropriate to blame the perpetrator. A perpetrator would *not* be morally responsible for a war crime if she had an excuse that made blaming responses illegitimate.[13]

It may seem obvious that if a soldier knowingly, and in the absence of direct coercive pressure, commits a war crime, then he is morally responsible and blameworthy for doing so. But as we saw above, Doris and Murphy argue that our default presumption should be that perpetrators of war crimes are *not* morally responsible for their conduct. This conclusion follows from their contention that because soldiers are exposed to certain situational pressures, they "manifest cognitive impairments that profoundly degrade their capacity for moral judgment, and such impairments . . . preclude the attribution of moral responsibility" (2007, 26).

We have already conceded Doris and Murphy's point that the extreme pressures of active combat can undermine a soldier's moral responsibility for his behavior—in such circumstances, a soldier may be too confused and frightened for his behavior to be attributable to him in a way that makes moral blame appropriate. However, this concession does not address Doris and Murphy's larger claim that factors that are generally present in military life and training undermine soldiers' moral responsibility even when they are not in combat. These distal situational pressures are supposed to undermine moral responsibility because they impair moral functioning, and they impair moral functioning by encouraging soldiers to believe that morally objectionable behavior is in fact permissible or even required (2007, 39). When, due to his military training, a soldier comes to believe that an activity such as torture is permissible, it is difficult for him to grasp the moral considerations that speak

[13] There are other ways of conceptualizing moral responsibility, but many philosophers accept an account along roughly the lines just described. Strawson (1962) is the classic source for this influential perspective; see Wallace (1996); and Watson (2004) for important discussions of, and elaborations upon, Strawson's basic themes.

against torture. But in this case—as Doris and Murphy might argue—isn't it unfair to blame a soldier for his bad behavior and to target him with negative emotional responses like resentment? After all, if a soldier, after undergoing training and enculturation processes over which he has little control, comes to believe that torture is permissible, then perhaps it is unreasonable to expect him to refuse to use torture. And we might think that it is only fair to blame a person when it is reasonable to expect that person to refrain from the actions for which he is blamed.

Doris and Murphy offer the most explicit and detailed argument of which we are aware for the blamelessness of those who commit war crimes. However, many other theorists accept the view that if, through no fault of her own, a wrongdoer regards her behavior as permissible, then it is unfair to blame her for her wrongdoing. In contrast to Doris and Murphy, these other authors generally do not appeal to the situationist experimental literature, but they do focus on the way that environmental factors can constrain one's moral outlook. For example, Susan Wolf claims that,

> we give less than full responsibility to persons who, though acting badly, act in ways that are strongly encouraged by their societies—the slaveowners of the 1850s, the Nazis of the 1930s, and many male chauvinists of our fathers' generation, for example. (1987, 56–57)

What excuses these wrongdoers is that they "falsely believe that the ways in which they are acting are morally acceptable" (1987, 57). Possession of these false moral beliefs tends to excuse the wrongdoers in question because the values that inspired these beliefs "may have been inevitable, given the social circumstances in which they developed" (1987, 57). Wolf concludes that "[i]f we think that the agents could not help but be mistaken about their values, we do not blame them for the actions those values inspired" (1987, 57). Philosophers who accept a view like Wolf's, and who also agree with Doris and Murphy about the impact of military training and martial culture on soldiers' moral capacities, should accept the conclusion that those who commit war crimes are often not open to the full force of moral blame.[14]

4. Blaming Perpetrators

The argument presented in section 3 is important and challenging, but we believe that there is little reason to accept its conclusion. On our view, military personnel may be blameworthy for their crimes even if it would be

[14] For defenses of views related to Wolf's, see Zimmerman (1997); Levy (2003); Rosen (2004); and Fricker (2010). For criticisms of such views, see Talbert (2013).

unreasonable to expect them to be aware of the moral status of their behavior and even if their behavior is informed by values and beliefs that they could not have avoided acquiring. Such perpetrators are potentially blameworthy because the fact that they regard their behavior as permissible is compatible with their behavior having the kind of interpersonal moral significance that makes blaming responses like resentment and indignation appropriate.

Emotional responses like resentment are reactions to the perception that one has been treated with contempt, ill will, or at least an objectionable sort of disregard.[15] Such responses are inapt—and a wrongdoer excused from blame—in cases where harmful behavior is in fact not expressive of contempt, ill will, or lack of regard. This is why we typically do not blame people who injure us only because they were, through no fault of their own, ignorant of the harmful consequences of their behavior, or when their harmful behavior was justified.[16] Similarly, we do not generally blame those whose harmful behavior was involuntary in certain ways (for example, because the agent was unconscious at the time of action, or because the behavior was the result of coercion, and so on).[17] In each of these cases of excuse, something blocks the otherwise natural inference that a person's harmful behavior shows that she lacks an appropriate degree of concern for others' welfare.

The fact that a perpetrator of war crimes believes that his behavior is permissible or justified does not provide him with any of the excuses just described. Obviously, the fact that a soldier has such a belief would not entail that his behavior is involuntary. And while a soldier who, for example, believes that torture is permissible may suffer from a form of ignorance, he is not ignorant of the consequences of his behavior. A torturer who thinks that he behaves permissibly is aware that he harms his victims (indeed, having that awareness is surely a necessary part of what it means to torture someone); he is only unaware that it is *wrong* to harm them. This *moral* ignorance—like that of the torturer interviewed by Martha Huggins who claimed to have acted within the bounds of his conscience—is fully compatible with the torturer having ill will—or at least a lack of appropriate regard—for his victims. Indeed, a plausible default assumption is that such a torturer's behavior is evidence that he

[15] Following Strawson, many philosophers reserve the term "indignation" for the third-personal analogue of resentment; in other words, it is the negative reactive attitude that we direct toward a wrongdoer on account of how he has treated a third party. We will speak mainly of resentment in what follows, but our position should be understand to apply to cases in which we blame a wrongdoer on behalf of his victims.

[16] The "through no fault of their own" caveat is important: it is often argued that if a person is at fault for her ignorance of the likely consequences of her behavior, then her ignorance will not excuse her from blame for these consequences. For an important discussion of these issues, see Holly Smith (1983).

[17] Angela Smith (2005) argues persuasively that our involuntary actions and states of mind are not always blameless; this is because some things that are not under our control (such as what we forget or fail to notice) can indicate our degree of concern for others.

implicitly denies that his victims have the sort of moral standing that would make torturing them illegitimate. If we reject the torturer's perspective on his victims' standing, then we should conclude that the torturer does not treat his victims with the degree of respect that we believe that they could reasonably demand, and that he is therefore an appropriate target of his victims' blame.

To take a somewhat more complicated example, consider the way doctors who served at Auschwitz rationalized their participation in the selection of prisoners for gassing and for medical experiments by reference to the moral goals of legitimate medical practice. One doctor stated: "Of course I am a doctor and I want to preserve life. And out of respect for human life, I would remove a gangrenous appendix from a diseased body. The Jew is a gangrenous appendix in the body of mankind" (Lifton 1986, 427). This doctor understood that his profession was founded on moral ideals such as respect for human life—in fact, he defended his actions by appealing to such ideals. But it is clear that his belief that Jews posed a threat to the health of mankind meant that he didn't see the principle of respect for human life as entailing respect for *Jewish* human life. This doctor was ignorant of important moral facts, but his ignorance was clearly compatible with—indeed, we may even think of it as constituted by—disregard or contempt for the interests and welfare of Jews.

But perhaps the Auschwitz doctor can avoid blame by offering his belief that the Jews were like a diseased organ as a justification of his behavior. In general, if one behaves justifiably, then one is not open to moral blame, and even false beliefs can sometimes provide an agent with justification for her behavior. For example, if a police officer shoots a suspect who is threatening him with a gun, the police officer's action is justified even if it turns out that the gun was fake, so long as it was reasonable for the police officer to believe the gun was real and that the offender posed a deadly threat that could only be averted through the use of force.

We can read the doctor's attempted justification in different ways, but none of them is likely to be persuasive. We might, for example, interpret the doctor's comparison of the Jews to a gangrenous appendix as just an expression of his bias against Jews: as if he were inviting us to believe that the Jews deserved harsh treatment because they are contemptible. But this clearly begs the question—we would hardly accept this as a justification of the doctor's behavior unless we already shared his biased outlook on Jews.

But perhaps this is the wrong way to think about the doctor's attempted justification. Perhaps the doctor's position is like that of Karl Kretschmer, the German soldier discussed earlier, who thought that Jews were a mortal threat to Germany. On this reading, the doctor is not simply inviting us to share his contempt for Jews; he is (like Kretschmer) justifying his behavior by reference to the supposed threat that the Jews posed. Now we might disagree with Kretschmer and the doctor as to whether the Jews really were a threat, but *if* they came by these false beliefs in a reasonable fashion, then we might think

that they are not open to blame for their bad behavior. Of course, there's good reason to doubt that Kretschmer and the doctor came by their beliefs reasonably. It seems to us much more likely that these two suffered from what Nomy Arpaly calls "motivated irrationality" (2003, 101–111). They were not making an "honest mistake" about the threat posed by Jews. Rather, their beliefs—which would have been transparently false to a person of good will—were nurtured and held in place by the kinds of objectionable attitudes and ill will that make blaming responses appropriate. We suspect, for example, that Kretschmer and the Auschwitz doctor did not fairly assess evidence that conflicted with their beliefs: they were, as Arpaly puts it, "ready to believe anything about Jews, as long as it was bad" (2003, 103).

So far, we have assumed that moral excuses function primarily by demonstrating that an agent's action was not expressive of objectionable attitudes like contempt. Thus, on our view, if a Nazi doctor treats Jews with contempt, it is for that very reason unlikely that he will have an excuse that allows him to avoid blame. However, this perspective on excuses can be questioned. Erin Kelly, for example, has argued that approaches like ours are too narrow since, as Kelly sees it, some agents whose behavior expresses contempt or ill will are not properly open to blame (2013). On Kelly's view, "excuses function by undermining what normally are reasonable expectations about how a person should be motivated" (2013, 256). According to this approach, when it comes to evaluating a wrongdoer's blameworthiness, we should consider to what degree it is reasonable to expect the wrongdoer to have behaved differently, and this will often depend on how difficult it would have been for the wrongdoer to avoid her bad behavior. For Kelly, then, some agents who definitely treat others with contempt are still not blameworthy because of how difficult it would have been for them to behave differently.

There is certainly something appealing about the idea that it is unfair to blame a person for an action when she would have had great difficulty acting otherwise. However, we suspect that this perspective is most plausible in those cases in which the very thing that makes it difficult for the agent to act otherwise also undermines the attribution to that agent of the attitudes to which blame responds. On our view, what is really doing the excusing work in these cases is not simply the agent's difficulty in acting otherwise, but the fact that the agent's actions do not express blame-grounding attitudes, such as contempt. For example, suppose that a soldier is ordered to shoot unarmed civilians and finds it very difficult to disobey because he is physically forced to carry out the order or because he knows that he will be shot if he refuses to obey. These explanations for why the soldier finds it difficult to disobey the order make it much less plausible that his behavior can be explained in terms of his attitudes toward those affected by his behavior. Even if he obeys the order and shoots the civilians, it is still possible that the soldier had no ill will toward them. Depending on the details of the case, on our view the soldier

might be excused for his behavior, but the reason why he might be excused would *not* be because he found it hard to avoid obeying the order, but because of the absence of ill will.

By contrast, imagine a soldier who has difficulty refraining from killing civilians because of the contempt he feels for them—he just thinks enemy civilians are scum who deserve to be killed. In this case, we would not be at all inclined to excuse the soldier's behavior regardless of how difficult he might have found it to act otherwise.

We believe, then, that some agents who have difficulty avoiding wrongdoing are blameworthy—namely, those agents whose difficulty does nothing to call into question their contempt or disregard for others. A perpetrator's belief that bad behavior is permissible may make it difficult for him to avoid bad behavior, but that kind of non-exculpatory difficulty gives us no reason to think that his behavior doesn't express contempt or disregard for others.

Another potential objection to our view involves the assumption that the point of blame is to initiate moral communication with a wrongdoer. For example, we might believe that the purpose of expressions of blame is to elicit a sincere apology from a wrongdoer and to inspire moral recognition on his part. On this view, any conditions that undermine a wrongdoer's ability to take part in this sort of moral "conversation" will also tend to make him an inappropriate target for the responses involved in moral blame. It has been argued on these grounds that psychopaths, for example, are exempt from blame because of their inability to engage with moral reasons and in moral conversations in the way that non-psychopaths do.[18] We do not believe that this objection tells us much about the case of most perpetrators of war crimes. For one thing, while we grant that military training can severely limit perpetrators' ability to recognize certain moral considerations and claims, this does not make them unfit for or incapable of moral reflection and conversation. The average perpetrator is not a psychopath: he does not lack a moral point of view; rather, his inability to respond to certain moral considerations is a function of his commitment to a flawed moral point of view. This is evident from the words of the perpetrators discussed earlier, all of whom rationalized and justified their actions by reference to moral considerations and moral reasons.

Still, once a perpetrator is in the grip of a flawed moral perspective, he is not likely to respond to moral blame in the way that we might wish: he is unlikely to apologize, to alter his behavior, and so on. But we do not believe that this makes blame inappropriate. On our view, the purpose of blame is not exhausted by the possibility of eliciting certain responses from a wrongdoer. Even if a victim cannot hope that her blame will inspire moral recognition

[18] See Shoemaker (2011); and Watson (2011).

in others, it can still be a means by which she stands up for herself and protests the treatment that she received.[19] The validity of such a protest does not depend on anyone being converted by it, and insofar as we agree that a victim has not been treated with due regard, we should view her moral protest, expressed through her blaming responses, as appropriate.

Our view is that moral blame is fundamentally a response to the objectionable attitudes that a wrongdoer can have toward those harmed by his actions. Military training and martial culture may impair the sensitivity of military personnel to moral reasons and incline them toward the view that various forms of objectionable behavior are permissible. These factors may in turn make it difficult for military personnel to refrain from committing war crimes and may make them impervious to the moral entreaties of their victims. However, none of these things makes it impossible for military personnel to guide their behavior by their own judgments about how they ought to behave; thus, none of these things entails that the behavior of military personnel cannot express the objectionable attitudes to which blame properly responds.

5. Luck and the Victim's Perspective

The role of luck in explaining perpetrators' behavior can raise concerns about their moral responsibility. In *The Lucifer Effect*, Philip Zimbardo cites a number of situational factors to explain why US Army reservists abused prisoners at the Abu Ghraib prison in Iraq: the reservists were untrained for their specific tasks, they were under pressure from military and civilian authorities to help produce intelligence, and they lived in a stressful and often dangerous environment. As a result, their conception of what was acceptable behavior began to shift from what it had been in civilian life. The following statement from Ken Davis, a military police officer at Abu Ghraib, sums up the situation:

> We were never trained to be guards. The higher-ups said, 'Use your imagination. Break them. We want them broke by the time we come back.' As soon as we'd have prisoners come in, sandbags instantly on their heads. They would flexicuff 'em; throw 'em down to the ground; some would be stripped. It was told to all of us, they're nothing but dogs [. . .]. So you start breeding that picture to people, then all of a sudden, you start looking at these people as less then human, and you start doing things to 'em that you would never dream of. And that's where it got scary. (Zimbardo 2008, 352)

[19] For more on the relationship between protest and blame, see Talbert (2012); and Angela Smith (2013).

As with the students in his Stanford prison experiment, Zimbardo doesn't think that the behavior of the soldiers at Abu Ghraib is explained by their possession of antecedent dispositions toward cruelty. The behavior was caused, rather, by the kinds of factors described in the quotation from Ken Davis and by what Zimbardo elsewhere describes as the "intrinsically pathological" prison environment (Haney, Banks, & Zimbardo 1973, 89).

But how did the reservists end up in these circumstances? This was largely a matter of luck. The reservists voluntarily enlisted for military service, but prior to their enlistment they had no way of knowing that they would end up at Abu Ghraib and that they would be subjected to the situational pressures that Zimbardo describes. Similarly, as Zimbardo notes, "[c]hance ruled" in determining the outcomes for the subjects in his Stanford experiment (2008, 206). It was "by virtue of random assignment" that "these good people were . . . assigned to the role of guard or prisoner" and subjected to pressures that elicited either objectionable or unobjectionable behavior (2008, 206).

When we focus on soldiers' lack of control over their exposure to situational pressures, it may seem unfair to hold them morally responsible for the behavior fostered by these pressures.[20] This worry can also be put in terms of the CAPS approach that we outlined earlier. Soldiers have little control over the external forces that help to shape their acquisition of CAPS traits, so, again, it may seem unfair to hold them morally responsible for the behavior that these traits inform.

Our response to these concerns echoes the reasoning we employed earlier. If someone is unlucky enough to have been exposed to pressures that make him unaware of the consequences of his behavior, or unable to control his behavior in light of his judgments about reasons, then he is excused from blame because his behavior is not likely to express the sort of contempt or disregard to which blame responds. However, the perpetrators we have been considering are not unlucky in this way; they remain fully capable of expressing contempt or disregard for others through their behavior.

But perhaps considerations about luck make it unfair to blame perpetrators even if they are capable of treating others with contempt. An argument for this conclusion might go like this: If luck plays a sufficiently large role in explaining why a perpetrator has bad moral values (or bad dispositions), then he is not at fault for possessing these values, and if he is not at fault for possessing these values, then it is unfair to blame him for the behavior that these values inspire. We do not accept this argument. Imagine a committed Nazi soldier—perhaps one trained from a young age in the *Hitlerjugend*—who regards Jews as vermin. Obviously, this soldier has objectionable attitudes toward Jews. However, it might plausibly be said that he is not at fault for having these attitudes: perhaps

[20] For a detailed account of how luck interacts with moral responsibility, see Levy (2011).

the fault lies with the institution of the *Hitlerjugend* and with those responsible for it. On our view, this hypothetical lack of fault for his objectionable attitudes would not exempt the soldier from blame. What is required for blameworthiness is that the soldier's behavior is informed by his values and judgments in such a way that it has (or could have) blame-relevant significance for those affected by his behavior. And it is possible for the soldier's behavior to have this sort of significance even if unchosen external influences shaped him into the sort of person that he is.

Still, to some it may seem unduly harsh to blame perpetrators given that many perpetrators are also often victims.[21] It's worth bearing in mind, then, what it would mean to deny that perpetrators are blameworthy. Consider Doris and Murphy's observation that perpetrators are sometimes trained to use sexual violence against women. Suppose that a woman was raped by a perpetrator who acted freely and was fully aware of the woman's distress and of the effects of his actions on her. So far, the perpetrator does not seem to have an excuse. But now suppose that the perpetrator was trained in such a way that—through no fault of his own—he came to regard sexual violence against women as a legitimate way of achieving political or military goals. Should we now conclude that the perpetrator is excused and that blame would be inappropriate? If we accept this conclusion, then we must say not only that it is inappropriate for *us* to blame the rapist, but that it is also inappropriate for his *victim* to blame him. We find this conclusion very unpalatable.

On our view, the perpetrator's (perhaps unavoidable and nonculpable) belief that he behaves permissibly does not excuse him from blame because nothing about the content of his belief or how it was acquired makes it unreasonable for his victim to resent him. Her claim to have been treated with unjustifiable contempt and disregard is not undermined in any way by these facts about the perpetrator's beliefs.

A person harmed by another should withhold blame if the harmful behavior did not express ill will or disregard of his moral standing. This would be the case if the harm was accidental, justified, the product of coercion, and so on. But none of this applies to the above example. To say that the rape victim shouldn't blame her rapist asks her to view her rape as if it were the kind of treatment that does not warrant resentment, as if it did not constitute a direct and profound attack on her moral standing—in short, it asks her to accept her rapist's perspective. This asks the rape victim to accept an inaccurate interpretation of what happened to her and fails to adequately account for the wrong done to her. Thus, to suggest that moral blame is off the table fails to do justice

[21] For example, many torturers in military dictatorships were subjected to extremely harsh training methods involving brutalization and humiliation (Haritos-Fatouros 1988).

to the victim because it misdiagnoses the moral quality and interpersonal significance of the treatment to which she was subjected.

Again, we have argued that what is crucial for blame is the expressive significance of the actions for which a wrongdoer is blamed. A wrongdoer is excused from blame in those cases in which his behavior does not convey blame-grounding attitudes toward others. However, just because a wrongdoer is not at fault for having objectionable attitudes does not mean that these attitudes are not attributable to him and cannot be expressed through his behavior. If a soldier has the bad luck to be exposed to factors that corrupt his values, then this can make him an appropriate object of pity in certain respects, but this fact will not by itself make him an inappropriate target of blame.

Conclusion

In this chapter we have argued for two central claims. Firstly, we have argued that situationist accounts of war crimes such as that offered by Doris and Murphy are inadequate and fail to account for the ways in which the behavior of perpetrators of war crimes is informed and shaped by their unique goals, beliefs, and self-conceptions. Our dispositional theory of war crimes, based on the CAPS theory of personality, recognizes the impact of situational forces on individuals but also reveals how important it is to view perpetrators' behavior through the lens of their individual CAPS traits. Without such a perspective, important and distinctive differences between perpetrators are ignored. Secondly, we have argued that certain arguments in favor of excusing perpetrators of war crimes fail. The behavior of many perpetrators of war crimes reflects objectionable attitudes regarding their victims' moral standing, and negative reactive attitudes like resentment are appropriate responses to such behavior. In addition, to hold that perpetrators of war crimes should be excused is to deny that the victims of war crimes may reasonably feel resentment regarding the way they were treated. Such a consequence disrespects the experiences of victims and privileges the perpetrators' moral standpoint.

The claims that we defend in this chapter are separable. That is, one could accept our CAPS account of war crimes and still think that soldiers are often blameless for war crimes because one thinks, for example, that a soldier is excused if he believes that he acts permissibly. Or one might accept our account of the moral responsibility of perpetrators yet reject our dispositional account of war crimes. However, the two claims fit together naturally: our approach to moral responsibility argues for the centrality of wrongdoers' orientations toward others, and the dispositions hypothesized by CAPS theory play this role. Our application of CAPS theory emphasizes the ways in which perpetrators' behavior is typically not a product of internal states that have been "forced" upon them by external circumstances, but rather reflects their

individual moral beliefs, emotional states, self-conceptions, and judgments. Such perpetrators are well positioned to express through their behavior the kind of ill will, contempt, or disregard for others that make blaming responses appropriate.[22]

References

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. New York: Oxford University Press.

Bandura, Albert. 1999. "Moral Disengagement in the Perpetration of Inhumanities." *Perspectives of Social Psychology Review* 3: 193–209.

Blass, Thomas. 1993. "Psychological Perspectives on the Perpetrators of the Holocaust: The Role of Situational Pressures, Personal Dispositions, and Their Interactions." *Holocaust and Genocide Studies* 30: 30–50.

Doris, John, and Dominic Murphy. 2007. "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity." *Midwest Studies in Philosophy* 31: 25–55.

Fontana, Alan, and Robert Rosenheck. 1999. "A Model of War Zone Stressors and Posttraumatic Stress Disorder." *Journal of Traumatic Stress* 12: 111–126.

Frésard, Jean-Jacques. 2004. *The Roots of Behaviour in War*. Geneva, Switzerland: International Committee of the Red Cross.

Fricker, Miranda. 2010. "The Relativism of Blame and William's Relativism of Distance." *Proceedings of the Aristotelian Society. Supplementary Volume* 84: 151–177.

Gwynne Dyer, 2004. *War: The Lethal Custom* (New York: Carrol & Graf Publishers).

Haney, Craig, Curtis Banks, and Philip Zimbardo. 1973. "Interpersonal Dynamics of a Simulated Prison." *International Journal of Criminology and Penology* 1: 69–97.

Haritos-Fatouros, Mika. 1988. "The Official Torturer: A Learning Model for Obedience to the Authority of Violence." *Journal of Applied Social Psychology* 18: 1107–1120.

Harman, Gilbert. 2000. "The Non-Existence of Character Traits." *Proceedings of the Aristotelian Society* 100: 223–226.

Hinton, Alexander. 2004. *Why Did They Kill? Cambodia in the Shadow of Genocide*. Berkeley: University of California Press.

Huggins, Martha. 2000. "Legacies of Authoritarianism: Brazilian Torturers' and Murderers' Reformulation of Memory." *Latin American Perspectives* 27: 57–78.

Kelly, Erin I. 2013. "What Is an Excuse?" In D. Justin Coates and Neal A. Tognazzini (Eds.), *Blame: It's Nature and Norms* (pp. 244–262). New York: Oxford University Press.

Levy, Neil. 2003. "Cultural Membership and Moral Responsibility." *The Monist* 86: 145–163.

[22] This paper was made possible through the support of a grant from the Character Project at Wake Forest University and the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of The Character Project, Wake Forest University, or the John Templeton Foundation. We would like to thank Christian Miller for his helpful comments on this chapter, as well as audience members at The Ethics of War in the 21st Century conference, Stockholm, for their useful and insightful feedback.

Levy, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. New York: Oxford University Press.

Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. London: Tavistock Publications.

Miller, Christian. 2014. *Character and Moral Psychology*. Oxford: Oxford University Press.

Mischel, Walter. 1973. "Toward a Cognitive Social Learning Reconceptualization of Personality." *Psychological Review* 80: 252–283.

Mischel, Walter, and Yuichi Shoda. 1996. "Toward a Unified, Intra-individual Dynamic Conception of Personality." *Journal of Research in Personality* 30: 414–428.

Pauer-Studer, Herlinde, and David Velleman. 2011. "Distortions of Normativity." *Ethical Theory and Moral Practice* 14: 329–356.

Rosen, Gideon. 2004. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18: 295–313.

Ross, Lee, and Richard E. Nisbett. 2011. *The Person and the Situation: Perspectives of Social Psychology*. London: Pinter and Martin.

Roth, Paul. 2004. "Hearts of Darkness: 'Perpetrator History' and Why There Is No Why," *History of Human Sciences* 17: 211–251.

Russell, Daniel. 2009. *Practical Intelligence and the Virtues*. New York: Oxford University Press.

Shoemaker, David. 2011. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121: 602–632.

Smith, Angela. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115: 236–271.

Smith, Angela. 2013. "Moral Blame and Moral Protest." In D. Justin Coates and Neal A. Tognazzini (Eds.), *Blame: It's Nature and Norms* (pp. 27–48). New York: Oxford University Press.

Smith, Holly. 1983. "Culpable Ignorance." *Philosophical Review* 92: 543–571.

Snow, Nancy. 2009a. "How Ethical Theory Can Improve Practice: Lessons from Abu Ghraib." *Ethical Theory and Moral Practice* 12: 555–568.

Snow, Nancy. 2009b. *Virtue as Social Intelligence: An Empirically Grounded Theory*. New York: Routledge.

Strawson, P. F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.

Talbert, Matthew. 2012. "Moral Competence, Moral Blame, and Protest." *Journal of Ethics* 16: 89–109.

Talbert, Matthew. 2013. "Unwitting Wrongdoers and the Role of Moral Disagreement in Blame." In David Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility* (pp. 225–245). Oxford: Oxford University Press.

Wallace, R. Jay. 1996. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Watson, Gary. 2004. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." *Agency and Answerability: Selected Essays*. New York: Oxford University Press, pp. 219–259.

Watson, Gary. 2011. "The Trouble with Psychopaths." In R. Jay Wallace, Rahul Kamar, and Samuel Freeman (Eds.), *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon* (pp. 307–331). New York: Oxford University Press.

Wolf, Susan. 1987. "Sanity and the Metaphysics of Responsibility." In Ferdinand Schoeman (Ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 46–62). Cambridge: Cambridge University Press.

Wolfendale, Jessica. 2007. *Torture and the Military Profession*. Abingdon, Engl.: Palgrave Macmillan.

Zimbardo, Phillip. 2004. "A Situationist Perspective on the Psychology of Evil: Understanding How Good People Are Transformed into Perpetrators." In A. G. Miller (Ed.), *The Social Psychology of Good and Evil* (pp. 21–50). New York: Guildford Press.

Zimbardo, Phillip. 2008. *The Lucifer Effect: Understanding How Good People Turn Evil.* New York: Random House.

Zimmerman, Michael. 1997. "Moral Responsibility and Ignorance." *Ethics* 107: 410-26.

INDEX

abilism; *see also* virtue epistemology (VE)
  definition, 321
  fake barn cases, 323–324
  Gettier cases, 321–322, 325–326
Abraham (biblical character), 388
Abraham Maimuni, 530n
Abu Ghraib prisoner abuses, 686–687
Adam (biblical character), 382, 527–528, 580
Adams, R.M., 231n, 336–337, 633
admiration, 9, 252–254, 259–260, 265; *see also* exemplarism
  and conscientious reflection, 262–263
  different kinds of, 254–255
  and elevation, 257–259
  and emulation, 258
  of exemplars of moral goodness, 264–265
  moral theory based on, 251–252, 263
    advantages of, 265–266
    and theory of direct reference, 263–264
  for natural and acquired excellence, 255–257
    benefitting oneself and others, 258–259
  psychology of, 257–259
  trustworthiness of, 260–262
After Virtue (MacIntyre), 612–613
Akiva, Rabbi, 528
Alfano, M., 292–294, 296, 296n, 297, 299, 299n, 300, 306n
Algoe, S., 258–259
Allen, A., 32
Allport, G.W., 102, 131, 131n, 132, 493–494, 503
Alston, W., 538–539, 543–547
Altemeyer, B., 656
altruism, 7, 55, 61, 75, 397, 449, 451, 483n; *see also* Belief in the Existence of True Altruism (BETA)
  existence of, 76, 79, 88, 90–92
    empathy-altruism hypothesis, 77–78
    empirical approaches, 76–77
  lay belief in, 75, 80–88, 92
Anderson, G., 370n, 372, 372n, 375n
Annas, J., 352, 352n, 601n
Anscombe, G.E.M., 235
Aquinas. *See* Thomas Aquinas
Aquino, K., 54, 482, 482n–483n
Aristotelian virtue ethics, 5, 33, 35, 226, 275n, 352, 611n, 617, 621
Aristotle, 20n, 118, 192, 215, 215n, 218, 220n, 221, 225n, 227–228, 251, 273, 273n, 275, 315, 346, 351, 351n, 352, 352n, 353–354, 356n, 357, 357n, 359–360, 365, 379, 579, 584, 597, 599, 601n, 605, 613, 617, 621, 621n, 624, 633, 645–646
Arnett, J.J., 491
Arpaly, N., 684
Asch, S., 59–61, 103, 294, 301, 652–654
Ashton, M., 159
Auerbach, E., 579n
Augustine, 568, 568n
Austen, J., 535
authority, obedience to, 14, 652; *see also* obedience experiments
  individual differences in, 654–656, 660–665
  role of authority figures, 665
  role of intelligence, 664, 666–667

Baehr, J., 291, 306–307, 309n
Bales, K.L., 611
Bandura, A., 677
Banerjee, R., 478, 480
Banks, K., 423
Barker, R., 636
Baron, R., 620–622, 623n
Bartz, J., 611
Batson, C.D., 54, 76–78, 82, 132, 175, 175n, 176–178, 178n, 179, 179n, 180, 180n, 181–184, 431, 653
Baumeister, R.F., 394, 398
Baumgartner, T., 620
Beaman, A., 37
Beauvois, J-L., 659
Belief in the Existence of True Altruism (BETA)
  conclusion, 91–92
  descriptive statistics of, 85–86
  exploratory and confirmatory factor analyses, 84–85
  instructions, 83–84
  and moral character, 87–91
  and other lay beliefs, 86–87
  scale items, 84
Bem, D., 32
Ben Sira, 524
beneficence, 118, 118n, 220–230

Benet-Martinez, V., 495
Bernano, G., 364
Big Five character traits, 3, 12, 48, 54, 57, 102, 136, 150, 159, 492, 494–506, 509–510, 624, 637, 653, 661, 666
Blasi, A., 63, 394, 394n
Bleidorn, W., 54
Bloom, P., 206
Boerger, A., 481
Bourdieu, P., 599, 600n
brain, role in ethics, 111, 408, 605–609, 609n, 610, 612, 614–615, 617n, 618, 624; *see also* neurobiological accounts of character traits; neurobiological accounts of virtue ethics; oxytocin
Brainerd, D., 561
Brambilla, M., 104–105
Brandt, R., 28
Brooks, P., 188
Brothers Karamazov, The, 373
Bucar, E.M., 600
bullying, 471–472
Burger, J.M., 656–657
Burns, A., 481
Burns, M., 453, 458
Buss, D., 24

Cain (biblical character), 382
Caldwell, R., 543n
Calhoun, C., 533
Calvin, J., 556–559
Campbell, A., 36
CAPS (cognitive-affective personality system), 606, 615, 675–677, 687, 689
Caputi, M., 478, 485
Caravita, S.C.S., 479
Carlo, G., 473
Carlson, E.N., 57–59
Catherine of Siena, Saint, 254
Cawley, M.J., 495–496
character, 41, 41n, 43, 131n, 132; *see also* consistency of character traits; Loving Mum and Tough Career Girl—How Do Role Settings Affect Our Personality and Character Traits?; moral behavior; moral character; personality; self-control; situationism; virtues
  affecting ethical value, 233
  changing of, 412
  cognitive approach in studying, 493
  consistency of, 41, 131–134
  definition, 41–42
  density distribution model of, 136–137
  development of, 36, 463
  in emerging adulthood, 515–517
  formation, by grace, 538
  as moral character, 41n
  origins, 445
  personal values conception, 233–234
  vs personality, 131n
  personality as basis for, 43
  in philosophy, 4–5, 192–193
  in psychology, 2, 41, 131–133, 493–494
  in theology, 5–6
character development in the school years, 11–12
  aggressive behavior, 469
  bullying, 471–472
  importance of, 467–468
  individual differences, 468–470, 472, 478–480
  knowledge needed for moral development, 472
  moral agency, 473–475
  moral development, 472–474
  moral identity, 473
  peer relationships, 469–471, 478–480
  social acceptance, 468
  social prominence, 468
  and theory of mind (ToM), 472–473, 476–478
    behavior toward peers, 478–480
    moral behavior, 481–485
    moral motivitation, 480–481
    social constructivist view, 478–479
    social goals perspective, 479–480
Character Project (Wake Forest University), 1
character prospection, 492, 504, 512–516
character strivings, 500–501, 503–504
  in emerging adulthood, 490
    assessment of, 504–508
    relevance of, 491–493
    stability of, 508–509
character traits, 35, 503–504; *see also* Aristotelian virtue ethics; Big Five character traits; consistency of character traits; dispositional view; good character traits; personality traits; person-situation debate; stability of character traits
  accuracy of people's understanding of, 160
  and behavior, 20, 114, 614
  as behavior-forming dispositions, 674–675
  and bodily actions, 20
  categorial traits, 29
  causal determinants of intra- and inter-individual differences in moral behavior, 52–53
  causal role in explaining and predicting behavior, 613–614
  changing empathy character traits, 428–431
  as cluster of dispositional beliefs and desires, 613–614

cognitive-affective personality system (CAPS), 675–676
commonly shared characteristics, 43–44
continuous traits, 29
contributions of, 32–33
degrees, 29
dispositional view, 20–21
domain-generality, 394
domain-specificity, 395
in emerging adulthood, 490
assessment of, 504–508
relevance of, 491–493
stability of, 508–509
existence of, 34–35
global, 31, 35
innate, 206
local, 31, 35
mental states, 19–20
minimal thresholds, 27–29
mixed, 35, 622–623
moral self-attributes, 501–504
neurobiological accounts of, 14, 55, 606–607, 616, 619, 624–625
normativity, 33–34, 165–166
ontological grounding of, 614–615
as patterns, 24
and rational will power, 207–208
role of emotions, 193–209
social-cognitivist view, 675
sustaining of, by emotions
conditioning theory, 200–201
countering axiological entropy, 203–205
extinction, 202–203
prevention of axiological entropy, 198–200
two- and three-dimensional framework, 112–114
unconditionally good, 120
variability, 139–146
without emotions, 207
cheating, 168–173
Chisholm, R., 342
Christian ethics; *see also* Reformed Protestant ethics
and Stoic Ethics, 560–565
and virtue ethics, 5, 357–359, 361–364
Christian liturgy, 13, 572
communal nature of, 588
comparative self-address, 581–583
fictive actions, 575–576
immersing oneself in liturgical script, 578–579
indexical appropriation, 581, 583–585
liturgical reenactment, 572–576, 578–579, 582–583, 585–587
make-believe behavior, 576–577
and moral commitment, 587–588
narrative recasting, 580, 583
and normative transformation, 587–588
self-reflexive character of liturgical script, 580
Christian values. *See* humility
Christian virtues. *See* humility; virtue ethics
Chrysippus, 200n
Churchland, P., 610
Cicero, 215, 215n, 562n
civility, 533–534
Clinton, B., 129–130, 146
Clore, G.L., 197
cognitive-affective personality system (CAPS), 606, 615, 675–677, 687, 689
Cohen, J., 514
Confucius, 255, 257, 265
consequentialist virtues, 272–274, 281–282
consistency of character traits, 31–32, 41, 49–50, 131–134, 631; *see also* stability of character traits
cross-situational, 27, 30, 622
density distribution model, 49–51
within-person variablity, 49–50
and stability, 30n
and the study of character, 41
Conway, J.M., 54
Copelstone, F., 351n
Courbet, D., 659
Craik, K., 24, 636
Crisp, O., 560
Crisp, R., 233, 241
Crockett, M., 610
Crowne, D.P., 656
Cuneo, T.D., 578n
Curie, M.S., 254
Cushman, F., 326

Dalton, D., 186
Damasio, A., 207n
Dambrun, M., 658–659
Darley, J.M., 45, 54, 132, 653
Davis, K., 686–687
Day of Atonement, 371, 374–376, 378, 382
de Sousa, R., 202n, 204n
deferring, 271–272, 274
deification, 547
descriptive and normative statements, 33–34
desiring, moral theory based on, 251–252
DeYoung, C., 606
Di Blasio, P., 479
Diary of a Country Priest (Bernanos), 364
dignity, 338, 344, 534
dispositions
characteristic manifestations, 22
conditional view, 24–25

dispositions (*Cont.*)
definition, 24
vs inert properties, 23
made up of interrelated mental state dispositions, 25–26
and predictions, 23
stimulus events or conditions, 22
divine forgivingness in the Hebrew Bible, 368–369, 371, 374, 391
as annulment of obligation, 370
in the book of Jonah, 384–387
central biblical images of, 370
and divine affections, 379–380
vs divine atrocities, 386–388
and divine cognitions, 379–380
vs divine moral perfection, 389–390
God as healer, 372–375
God as lawgiver, 368
Hebraic moral identity, 373–374
hesed, 378, 380–381, 383–384, 386, 388
kipper, 370–371, 374–375, 378–379, 382–383
liturgical dimension, 374–375
nasa awon, 372, 375–378, 381–383
scapegoat, 374–376, 382
sin as pollutant to be expunged, 371, 371n
theological retrieval, 376–377
use of metaphors, 376–377
Domain-Specific Impulsivity Scale (DSIS), 402
Domain-Specific Impulsivity Scale revised (DSIS-7DS), 402–403
Donnellan, K., 265n
Doris, J., 35, 37, 46, 292, 294–295, 295n, 296–297, 302–303, 306n, 319, 605, 614, 621–622, 654, 672–673, 673n, 674, 676, 679–681, 688–689
Dostoevsky, F.M., 373
Dougherty, D.M., 56

Edwards, J., 538–539, 539n, 540, 540n, 541–542, 542n, 543, 543n, 545–548, 548n, 549, 549n, 550–551, 557–561, 564n, 567–568
Eisenberg, N., 62, 416
Eisenberg-Berg, N., 63
Eleazar, Rabbi, 528
Elms, A.C., 656
Emmons, R.A., 501n
emotions and character traits, 193–209
empathy, 11
changing of, with mobile phones, 412
cognitive, 419, 422
definition, 413
emotional, 418–419, 422
impact of new technology on, 417
importance of, 413–414
and kindness, 413
malleability of, 416
prosocial behavior, 414, 419, 422–424
stability of, 415–416
and well-being and health, 415, 431
empathy-altruism hypothesis, 77–78
Epictetus, 559, 559n, 562–563, 563n, 564, 564n, 566, 569
Epistemic Authority (Zagzebski), 260n
epistemic situationism, 289
strong epistemic situationism about intellectual character, 292–294, 296
strong epistemic situationism about reliability, 292, 294–296
weak epistemic situationsm, 292
epistemic virtues
consequentialist view of, 275–276
deference, 274, 281, 285–286
compliance, 277–278
instilling of, 276–277
promoting of, 283–284
sources needed to promote, 280–281
deferring, 272, 284
procedural justice, 281, 284–286
components, 278–279, 279n
and consultation, 280
legitimacy of authorities, 278–279
and rule following, 279–280
Epstein, S., 134, 407, 641
Erasmus, 559, 559n
ethics. *See* Christian ethics; Reformed Protestant ethics; Stoic ethics; virtue ethics
etiquette in Judaism, 12, 522, 526–528, 531–536
Euclid, 358
eudaimonia, 251, 617
Eve (biblical character), 382
exemplarism, 9, 251, 264–266; *see also* admiration
exemplarity in Judaism, 522–524, 526–528

fairness
development in childhood, 446–448, 460
altruistic sharing task, 449–451
changing of fairness concerns, 451–453
characterological limitations, 462
individual differences in infancy, 461–462
violation-of-expectancy resource distribution task (RDT), 448–453, 455–457
distribution of goods and resources, 446–447, 460–461
individual differences in infancy, 454–455
infants vs parents, 455–457
integrating with competing social and moral motivations, 457–458
in infants, 458–460
nature of fairness concerns in infancy, 453
affecting affiliative behavior, 453–454

within-person variablity, 184
and perspective-taking, 180
principle of equality, 446
and self-awareness, 179
and self-interest, 181
and unfairness, 182–183
Falkowitz, S., 481
Fast, L.A., 655
Felps, W., 54
feminist virtue ethics, 591, 599–601
Flanagan, O., 208, 606
Fleeson, W., 50, 54, 134–136, 138, 396–397
Foot, P., 351n
forgivingness. *See* divine forgivingness in the Hebrew Bible
Fournier, M., 53
Francis, Saint, 265
Frankfurt, H., 381, 607n
Franklin, B., 118
Freeman, D., 54
Fretheim, T., 376, 390
Freud, S., 463
Fricker, E., 274n
Fries, A.B., 610n
Frimer, J.A., 64
Fritz Cates, D., 568
Funder, D.C., 396, 407, 655
Furr, R.M., 56–59

Gage, P., 624
Gallagher, P., 50
Gamaliel, Rabban, 528
Garrigou-Lagrange, R., 361n
Gasser, L., 480–481
Gauguin, P., 217, 217n
Gebauer, J.E., 54
Gert, B., 117, 118n
Gettier, E., 316, 325–326
al-Ghazali, 599n
Ghiberti, L., 255–256
Gini, G., 480
Goldberg, L.R., 502, 504
Goldman, A., 271, 276n, 285, 311n
good character traits
causal role in developing, 36
development and cultivation of, 36
role of religion, 36
secular ways of cultivating, 36–37
Goodall, J., 308
grace; *see also* infusion of grace
common, 540
external influence of, 539–540
as habit, 359–360
special or regenerative, 540
Thomas Aquinas, 356–357, 359–360
Graham, J., 117
Greco, J., 315
Greenspan, P., 204–205
Gregory, E., 568, 568n
Gregory XI, Pope, 254
Griffiths, P., 616, 616n

Haas, K., 656
Hagele, S., 54
Haidt, J., 117, 257, 257n, 258–260, 260n, 266, 407
halakhah, 523, 535–536
Hamlin, J.K., 61
Hardy, S.A., 473
Hargis, K., 77
Harman, G., 319, 605, 621–622, 654
Hartshorne, H., 41, 45, 48, 101, 101n, 132, 394–395, 494
Hazlett, A., 317
headscarves. *See* Islamic veiling in Indonesia; Muslim sartorial practices
Hebrew Bible. *See* divine forgivingness in the Hebrew Bible
Heller, D., 54
Hennig, K., 117
Herdt, J., 560
hesed, 378, 380–381, 383–384, 386, 388
HEXACO model of personality structure, 48, 102, 151–152, 159–160
hijab, 591–592, 594–598
Hinton, A., 679
Hippocrates, 605, 624
Hitler, A., 260–262
Hobbes, T., 75, 463
Hoffman, M.L., 77
Hogan, R., 491, 511, 515
Holland, C.H., 656
Holy Spirit. *See* indwelling of the Holy Spirit
Hookway, C., 291
Horney, K., 222
Hornstein, H.A., 81
Huggins, M., 679, 682
Hume, D., 315, 606n
humility
active module of (AH), 337, 343–344
objection that it is too demanding, 346–348
cognitive module of (CH), 337–340
objection that it is false, 344–346
objection that it is too demanding, 346–348
emotive module of (EH), 337, 340–343
objection that it is false, 344–346
objection that it is too demanding, 346–348
imitation of Christ, 336, 339, 341–342, 344–345
as social virtue, 333
in the New Testament, 334–336
Hurka, T., 216n, 218, 218n, 233–234, 237–243

Hursthouse, R., 365n
Hurtado, L., 336

Ibn Miskawayh, 596n
imitation of Christ, 336, 339, 341–342, 344–345
impulsivity, 393; *see also* self-control
  causal role of temptation, 403
  definition, 394, 394n
  domain-generality, 404
  domain-specific behavior, 397–398, 404–406, 408
    gender differences, 404, 406
    stability, 404
    studies on, 399–401
    subjective temptation and perceived harm, 398–400
    temporal discounting, 401
  Domain-Specific Impulsivity Scale (DSIS), 402
  Domain-Specific Impulsivity Scale revised (DSIS-7DS), 402–403
  as feature of psychopathy, 205
  in sexual behavior, 393, 406, 408
Indonesia. *See* Islamic veiling in Indonesia
indwelling of the Holy Spirit, 13, 539, 552
  Alston's model
    fiat model, 543–544
    integration with Edward's theological psychology, 545–546
    internality, 544
    interpersonal model, 543–544
    partial life sharing model, 544–545
  and Christological sonship
    aspect of adoption, 546
    enjoyment and communion, 547
    filial unitive drive, 548–552
    status of children, 546
    union with God, 547
  individual aspect of, 539
infusion of grace, 13, 538–539, 541–543; *see also* grace
  basis for spiritual life, 541
  and filial obedience, 550–551
  as indwelling Spirit, 541
intellectual virtues
  reliabilist conception of, 290, 290n, 291
  responsibilist conception of, 290, 290n, 291–292, 298–302
Iran, Islamic veiling in, 598
Isaiah (biblical character), 370, 370n, 376
Isen, A., 293, 301
Islamic veiling, 590, 590n
Islamic veiling in Indonesia, 13, 591; *see also* Muslim sartorial practices
  as fashion statement, 595, 598–600
  fashion trends, 593
  headscarf advice literature, 594–595
  local dress codes, 593
  and modesty, 597–598
  as political statement, 600

James, W., 318–319
Jayawickreme, E., 54
Jefferson, T., 257–259
Jesus Christ, 265, 335–336, 338–344, 348, 371n, 388, 546–548, 548n, 549, 552, 557, 574–576, 578, 580, 588
Jewish liturgy, 374–375, 532
jilbab, 591, 593–595, 598, 600
John, O.P., 496
Johnson, J., 24
Jonah (biblical character), 382, 384–387, 387n
Joseph, C., 117
Joseph of Arimathea, 579
Joshua ben Levi, Rabbi, 528, 531
Judah, Rabbi, 532
Judaism; *see also* divine forgivingness in the Hebrew Bible
  character and virtue in, 523
  ethical literature in, 523
  etiquette and law, 531–532
  etiquette between law and ethics, 533–536
  etiquette in, 522, 526–528
  exemplarity in, 522–524, 526–528
  God as lawgiver, 368
  halakhah, 523, 535–536
  and Islam in contrast to Christianity on shaping character, 523
  law and training in virtue, 523–525
  moral identity, 373–374
  role of law in, 522–523, 523n, 524
  role of sage in, 528–530
  theology of character in, 522
justice. *See* fairness

Kamp, J., 187–188
Kant, I., 115, 181n, 228, 347, 377, 561n, 599, 599n
Karylowski, J., 77
Keizer, K., 653
Keller, M., 480–481
Kelly, E., 684
Killen, M., 474
King, L.A., 643
kipper, 370–371, 374–375, 378–379, 382–383
Klein, K., 423
Kohlberg, L., 61
Komar, J., 54
Koonz, C., 260
Kosfeld, M., 620

Krakowski, E., 530n
Kretschmer, K., 678–679, 683–684
Kripke, S.A., 263–264, 265n
Kupperman, J., 120
Kvanvig, J., 309, 309n, 310

Lackey, J., 271n
LaKind, E., 81
Lancaster, B., 204
Landau, F., 678–679
Lapsley, D.K., 54, 502
Lasky, B., 54, 502
Latané, B., 45
lay beliefs; *see also* Belief in the Existence of True Altruism (BETA)
  in altruism, 75, 81
  in the existence of God, 80
  in free will, 79–80
  in a just world, 79
  in moral goodness, 80–81
  naive Hobbesians, 83
  naive Rousseauians, 83
  in pure evil and pure good, 82–83
  review of literature, 78–79
  self-perceived altruism, 82
  in universal egoism, 82
Leach, C.W., 104–105
Lecce, S., 478
Lee, K., 54, 159
Lerner, M.J., 79
Lewinsky, M., 129
Lewis, C.S., 336
Life of David Brainerd, The (Edwards), 561
Lim, V.K., 54
liturgy. *See* Christian liturgy; Jewish liturgy
Locke, J., 463
Loopik, M.v., 528n
Louf, A., 363–364
Love's Knowledge (Nussbaum), 578–579
Loving Mum and Tough Career Girl—How Do Role Settings Affect Our Personality and Character Traits? 7, 137–140
Lucifer Effect, The (Zimbardo), 652, 686–687
Luther, M., 557, 560–561

MacIntyre, A., 248n, 612–613, 615
MacLeod, A.E., 600
Maghen, Z., 523, 523n, 524–525, 535
Maimonides, 528–529, 529n, 530, 530n, 531–532
Maner, J.K., 78
Manne, S., 81
Margalit, A., 371
Markovits, J., 233, 241
Marsden, G., 567
Marsh, D.M., 56
Martin, J., 533
Mary of Bethany, 574–575, 575n, 576, 578, 587–588
Mary Theotokos (mother of Jesus), 579
Mathias, C.W., 56
Matsuba, M.K., 63
Mauthe, K., 481
May, M.A., 41, 45, 48, 101, 101n, 132, 394–395, 494
McAdams, D.P., 496
McCrae, R.R., 496
Mehl, M.R., 637
Meilaender, G., 565
Melchers, M., 618n
mental state dispositions, 25–26, 26n, 28, 29n, 166, 166n, 167–168, 174, 413, 613, 623, 623n
Merleau-Ponty, M., 548n
Metzger, M.B., 186
Micah (biblical character), 380
Milgram, S., 652–666, 672–673
Milgrom, J., 375n
military values and war crimes, 677
Mill, J.S., 272–274, 347
Miller, A.G., 657
Miller, C., 54, 79, 413, 607, 613–615, 615n, 618–619, 621, 621n, 622–623, 623n, 624, 656
Mischel, W., 24, 35, 132, 134, 395–397, 407–408, 494, 615, 675
Mishne Torah (Maimonides), 528–530
Mixed Trait Model of Character Traits, 8
  cheating, 168–173
  continuum for honesty and dishonesty, 167, 169
  fairness
    within-person variablity, 184
    and perspective-taking, 180
    and self-awareness, 179
    and self-interest, 181
    and unfairness, 182–183
  key assumption, 165
  mental state dispositions, 167–170
  moral character and self-awareness, 179–180
  moral hypocrisy, 177–178
  motivation, 172, 174
  overview, 164–165
  within-person variablity, 174
  resource distribution, 164, 175–178
  theft, 164, 184–188
mobile phones; *see also* text messages
  as habit former, 412
  shaping character, 412, 432

moral behavior
biological explanations, 55
consistency of, 48–52, 130, 644
contingencies of, 53–54
density distribution model, 52
linguistic stability, 640–643
momentary stability, 640, 642
observational studies of, 636
Electronically Activated Recorder (EAR), 14, 638–645
and oxytocin, 14, 606, 606n, 607, 618n, 620, 625
person-situation debate, 130
rank-order stability, 640, 642
social-cognitive variables, 54–55
stability of, 643, 645
moral character
coherence of, 645
development in early childhood, 61–64
differences become apparent in adolescence and adulthood, 63–64
fundamental virtues of, 118–119
inferred from behavior in social interactions, 114–116
measurement of, 87–88
multiple meanings, 100–101
perceived underlying structure of, 116–118
in person perception and evaluation, 100–101, 105
archival study, 109
correlational study, 106–108
experimental study, 108–109
fundamental to identity, 111–112
person judged on weakest link, 109–111
and warmth, 112
personological basis of, 3–4
ranking of virtues, 119–120
in self-knowledge and other knowledge, 55–56
self-perception and other's perceptions, 56–57
in social cognition research
communion, 104
consistency, 101
existence, 102
sociability, 104–105
study of personality preferred, 102
trustworthiness, 104
warmth, 102–105, 105n
traits as lenses or filters, 59–60
two people's perception of a third person, 57–59
moral molecule (oxytocin), 607–609, 609n, 610–611
Moral Molecule: The Source of Love and Prosperity, The (Zak), 608–609, 609n
moral risks, 9, 216–218
beneficence and perfectionism, 220–230
and moral luck, 217–218
non-moral goods and virtue, 230–231
in virtue ethics, 218–220, 231
moral self-attributes, 512
connections with character traits and character strivings, 510–512
in emerging adulthood, 490
assessment of, 504–508
relevance of, 491–493
stability of, 508–509
Morgan, M.L., 368–369
Morimoto, A., 540
Moses (biblical character), 375n, 376, 527
Moskowitz, D.S., 53
Murphy, D., 672–673, 673n, 674, 676, 679–681, 688–689
Muslim sartorial practices, 13, 591; *see also* Islamic veiling in Indonesia
beauty, 598–599
bodily practices, 595–596
modesty, 597–598
political dimensions, 599–600
Qur'anic directives to veil, 13, 591–592, 592n, 595, 597–598
scriptural context, 591
virtues of women's dress, 591–593
virtuous lifestyle, 596–597
Mussen, P., 63
Mutiara, I., 594

narrative identity, 573, 576–577
narrative work
contribution to moral life, 579
immersing oneself in, 577–578
Narvaez, D., 54
nasa awon, 10, 369–370, 372, 374–378, 381–383
Nazi genocides, 678–679, 683–684
Nelson, C., 102
neurobiological accounts of character traits, 14, 55, 606–607, 616, 619, 624–625; *see also* brain, role in ethics; oxytocin
neurobiological accounts of virtue ethics, 606, 606n, 607, 611, 615–620, 624–625; *see also* brain, role in ethics; oxytocin
Newcomb, T.M., 132
Nichols, S., 111
Nicomachean Ethics, 251, 273n
Nietzsche, F., 379
Nisbett, R.E., 45
Noftle, E.E., 64, 397, 498
Norman, W.T., 497
normative and descriptive statements, 33–34
Nosek, B., 117
Nussbaum, M., 199, 200n, 202, 222, 260n, 563n, 566, 578–579

obedience experiments, 653–654; *see also* authority, obedience to
  contemporary, 656–660
  immersive virtual reality (IVR), 658
  use of virtual reality, 658–661
Obedience to Authority (Milgram), 655, 663
Oberlé, D., 659
O'Brien, E.J., 407
Odbert, H.S., 494
Olberding, A., 523–524
Olin, L., 292, 294–295, 295n, 296–297, 302–303, 306n
One Boy's Day (Barker & Wright), 636
oxytocin
  and Mixed Trait theory, 623, 623n, 624
  and moral behavior, 14, 606, 606n, 607, 618n, 620, 625
  as moral molecule, 607–609, 609n, 610–611
Ozer, D.J., 396, 495

Packer, D.J., 657, 662
Pagnin, A., 478
Park, N., 138
Park, Y., 474
Pauer-Studer, H., 678, 678n
Paul, Saint, 335–336, 338–340, 345, 348
Paunonen, S.V., 502n
Peirce, C.S., 315
Pennebaker, J.W., 643
perfectionism, 155, 220–230
personality
  as basis for character, 43
  cognitive-affective personality system (CAPS), 675–676
  coherence, 632–633, 645
  consistency and stability in behavior, 632, 635
  different approaches to, 43
  dispositions, 654
  neuroscience, 606
  observational studies of, 636
    Electronically Activated Recorder (EAR), 637–638
  in psychology, 2–3, 7
  renaissance of, 46
  self-reports, 634–636
Personality and Assessment (Mischel), 132, 395–396
personality traits
  biological explanations, 55
  and character traits, 19, 47, 605–606
  density distribution model, 134–136, 396
  domain-generality, 395–396
  domain-specificity, 396–398, 407–408
  evidence for, 46
  and moral behavior, 47–48
  within-person variablity, 407–408
  prediction of life outcomes, 47
person-situation debate, 133, 394, 396, 630–632; *see also* situationism
  obedience to authority, 652, 654, 656
Peter, Saint, 360
Peterson, C., 117, 137–138, 500, 505
Pettit, P., 341
Philo of Alexandria, 524–525, 535
Philosophos, 528
phronesis, 346–348, 352–355, 364, 616–617; *see also* practical wisdom
Pippin, R., 533
Pitts, R.C., 54, 501–502
Plato, 215n, 315, 573, 621n
Plotinus, 389
Polanski, R., 90
Politics and the Order of Love (Gregory), 568
Porter, J., 351n
post-traumatic stress disorder (PTSD), 202–203
Powell, A.A., 78
practical wisdom, 10, 205, 225, 346, 348, 352, 352n, 353–355, 362, 364–366, 535, 549, 617; *see also* phronesis
Pride and Prejudice, 535
Protestant ethics. *See* Reformed Protestant ethics
psychology of grace, 538
PTSD (post-traumatic stress disorder), 202–203
Putnam. H., 263, 265n

Quine, W.V., 326
Qur'anic directives to veil, 13, 591–592, 592n, 595, 597–598

Ramsey, P., 540n
Rania (fashion blogger), 594–595, 598, 600
Rashi, Rabbi, 530
Rav (Abba Arikha), 531–532
Rava (Abba ben Joseph bar Hama), 532
Reed, A., 54, 482, 482n–483n
Reformed Protestant ethics, 13
  divine and human agency, 555–560, 563–565, 569
  interrelation of human beings, 565–567
  love for universal being, 565, 567–569
  necessity and moral responsibility, 559–560
  recognition of own moral limitations, 561–562
Reid, T., 315
Rimé, B., 204
Roberts, R., 307–309, 309n, 310, 310n, 366, 373, 379, 492
Rogelberg, S.G., 54
Rosen, M., 534
Rosenberg, M., 80–81, 102–103
Rosenberg, S., 113
Rosenhan, D.L., 77

Ross, L., 45
Ross, W.D., 118
Rousseau, J.-J., 75, 463
Rushton, J.P., 36
Russell, D., 352n

saints, 366, 541–542, 543n, 544, 547
Salmivalli, C., 479
Salmon N., 265n
Salovey, P., 77
salvation, 13, 335, 555–556, 561–562, 564, 569
Saucier, D.A., 82–83
Saucier, G., 502, 504
Saxe, R., 618n
Schmidt, M.F., 61
Schopenhauer, A., 101n, 111
Schuessler, K.F., 81
Schwartz, S., 77, 181n
Search, The (C.P. Snow), 255
self-control, 11, 80, 151, 155–156, 393–394, 394n; *see also* impulsivity
self-other knowledge asymmetry (SOKA), 635
Seligman, M.E.P., 117, 137–138, 500, 505
Sellars, W., 315
Seneca, 564n
sexual behavior, impulsivity in, 393, 406
Shaw, B., 255–256
Sherif, M., 293–294, 301
Shimon bar Ba, 526
Shoda, Y., 396, 615, 675
situationism, 3, 35, 44, 132–133, 671n; *see also* person-situation debate
  consequences of, for academic research, 45–46
  ethical, 288–289
  evidence for, 45
  experimental research supporting, 671
Slater, M., 658–660
Slora, K., 185–186
Smith, A., 77, 682n
Snow, C.P., 255–256, 259
Snow, N., 54, 675
Socha, L., 254
social virtues. *See* humility
SOKA (self-other knowledge asymmetry), 635
soldiers. *See* war crimes
Sommerville, J.A., 61
Sorabji, R., 563n, 567
Sosa, E., 315
Spiegel, J., 345n, 346
Spitzer, E., 393, 398–399, 408
stability of character traits, 29, 30n, 31, 622, 631; *see also* consistency of character traits
  aggregate trait stability, 29
  single-situation trait stability, 29
Stich, S., 77
Stohr, K., 228, 534–535
Stoic ethics, 555
  and Christian ethics, 560–565
  moral obligations toward universal humanity, 566–567, 569
Strange, S.K., 563
Strawson, P.F., 682n
Strimple, R., 335n
Strohminger, N., 111
Sutton, J., 479
Swanton, C., 222–223
Sweet Smell of Success (movie), 204

tapeinophrosune. *See* humility
Tarlo, E., 593n
text messages, 418, 420–421; *see also* mobile phones
  changing empathy character traits, 418–420, 426–428
  as empathy-building intervention, 428–430
  as intervention tools, 417–418
  study design and participants, 422–425
    data analysis, 426
    engagement metrics, 425–426
    outcome measures, 420–421
theft, 164, 184–187
theosis, 547
Thomas Aquinas, 77, 351, 351n, 352, 352n, 355–356, 356n, 357, 357n, 358–361, 361n, 362–363, 363n, 364–366, 579
Thomley, J., 622
true altruism. *See* altruism; Belief in the Existence of True Altruism (BETA)
Tyler, T., 278–279, 279n–280n

Utilitarianism (Mill), 272

Valadez, S., 255
Values in Action Inventory of Strengths (VIA-IS), 498, 500–501, 505
Values in Action (VIA) classification of character strengths, 137–138, 497–500, 505
Vatine, E., 658–659
Vazire, S., 635
veiling. *See* Islamic veiling in Indonesia; Muslim sartorial practices
Velleman, D., 678, 678n
Verbrugge, V., 334
Viechtbauer, W., 492
virtue epistemology (VE), 9–10; *see also* abilism; intellectual virtues
  benefits of, 315–316
  challenged by concept of belief, 318–319
  challenged by concept of reliability, 319–321

challenged by concept of truth, 317–318
reliabilist conception of, 302–305
strong autonomous responsibilism, 305, 309–310
weak autonomous responsibilism, 305–309
weak conservative responsibilism, 305–307
situationist arguments against, 296–302, 310–311
empirical inadequacy, 289
normative paucity, 289
skepticism, 289
virtue ethics; *see also* Aristotelian virtue ethics; phronesis; practical wisdom
active life as greatest of goods, 215
acts of infused virtue, 362–363
acts of natural virtue, 359
Aristotle, 351–352, 352n, 353–355, 617
and character traits, 288
conceptual fragmentation, 616–618
contemporary Christian, 363–364
core thesis, 233
dispositions vs occurrent mental states and behavior, 234–235
personal character, 245
personal values and personal character, 243–244
personal values conception, 237–242, 248–249
reliable trait conception of character, 236–237, 240–242, 248–249
eliminativism and epiphenomenalism, 615–616, 618
and empirical content, 611–612
explanatory project in, 612
first moral principles, 357–359
moral risks, 215
natural flourishing, 356–359
neurobiological accounts of, 606, 606n, 607, 611, 615–620, 624–625
pureley aspirational project in, 612
role of motivation and deliberation in moral behavior, 614
self-sufficiency, 351–352, 354
situationist challenge, 275n, 620–622, 630–633
supernatural flourishing, 356, 359–361
Thomas Aquinas
challenges to, 365–366
and contemporary accounts, 364
gifts, 362–363, 363n
marked by dependency, 351–352, 355, 361
practical wisdom, 352n, 364–365
role of Holy Spirit, 361–363, 363n
virtues, 35–36; *see also* character; Christian virtues; consequentialist virtues; humility; self-control
causal role in developing, 36
dispositional view, 274
fundamental of moral character, 118–119
instability over time, 275
and Islamic veil, 590
modularity of, 336–337
normative standards for, 165–166
ranking of, 119–120
and rational will power, 207–208
role of religion, 36
secular ways of cultivating, 37
thin and thick conception of, 222
unconditionally good, 120
Vivekananthan, P.S., 102
von Wright, G.H., 207

Wainryb, C., 475, 475n
Walker, L.J., 63–64, 117, 501–502
Walter, N., 618n
Walton, K.E., 492
Walzer, M., 216
Wang, Z., 610n
war crimes, 14–15
Abu Ghraib prisoner abuses, 686–687
cognitive-affective personality system (CAPS), 676–677, 689
dispositional view, 674, 679, 689
meaning making of perpetrators, 678–679
moral blame of perpetrators, 681–689
moral disengagement, 677
moral ignorance of perpetrators, 682–683
moral responsibility of perpetrators, 680–689
Nazi genocides, 678–679, 683–684
Nazi physicians in Auschwitz, 683–684
perceived as consistent with military values, 677
person-situation debate in explaining, 671
situationist accounts of, 672–674, 679–681, 689
soldiers' moral responsibility for, 672–674
soldiers' normative competence compromised, 674
Weber, M., 600n
Webster, R.J., 82–83
WECT. *See* Work Experiences and Character Traits (WECT)
Whittaker, J., 338
Wiggins, J., 24
Williams, B., 217, 217n
Wimbush, J., 186

Wolf, S., 681
Wolterstorff, N., 371n, 389
Wood, J., 307–309, 309n, 310, 310n
Work Experiences and Character Traits (WECT), 8, 150–153, 161
  consistency of character traits, 158–159
  evidence for the existence of character traits, 158–159
  and moral character, 153–158
  self-other agreement, 160
Wright, H., 636
Wrightsman, L.S., 81

Yohanan, Rabbi, 526–527, 527n
Yom Kippur, 371, 374–376, 378, 382
Young, L., 326, 617n–618n

Zadra, J.R., 197
Zagzebski, L., 290–291, 291n, 294–296, 297n, 298, 298n, 299, 299n, 300, 307, 315, 346–347, 353n, 365n, 523–524
Zak, P., 608–609, 609n, 611, 620–621, 623n, 624
Zeno, 562
Zimbardo, P., 652–654, 672–673, 686–687
Zuroff, D., 53